Mass Screening Registry
Finnish Cancer Registry
Helsinki, Finland

Department of Obstetrics and Gynaecology
Helsinki University Central Hospital, Finland

Department of Public Health
Hjelt Institute
University of Helsinki, Finland

CASE–CONTROL STUDIES FOR THE EVALUATION OF PERFORMANCE AND AGE-SPECIFIC OUTCOME OF ORGANISED CERVICAL CANCER SCREENING

Stefan Lönnberg

ACADEMIC DISSERTATION

To be presented by permission of the Medical Faculty of the University of Helsinki
for public examination in the Biomedicum lecture hall 3,
Haartmaninkatu 8, Helsinki, on Friday, 9 November 2012, at 12 noon.

Helsinki 2012

Supervised by

Docent Ahti Anttila
Mass Screening Registry
Finnish Cancer Registry

Docent Pekka Nieminen
Department of Obstetrics and Gynaecology
Helsinki University Central Hospital

Reviewed by

Professor Johanna Mäenpää
Department of Obstetrics and Gynaecology
University of Tampere and Tampere University Hospital

Docent Simopekka Vänskä
National Institute for Health and Welfare

Official opponent

Professor Joakim Dillner
Department of Laboratory Medicine and
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet

ISBN 978-952-10-8344-0 (paperback)
ISBN 978-952-10-8345-7 (PDF)
http://ethesis.helsinki.fi/

Cover illustration:
'Virukset tekevät syöpää' - 'Virions making cancer',
impression of oncogenic viral particles infecting host cells,
Erik Lönnberg (4)

Unigrafia Oy
Helsinki 2012

"The ultimate goal is to manage quality. But you cannot manage it until you have a way to measure it, and you cannot measure it until you can monitor it."

Florence Nightingale (1820 –1910)

TABLE OF CONTENTS

1 ABSTRACT

Screening for cervical cancer has the potential to reduce cancer incidence and cancer death by over 80%, but this level of effect requires an organised programme with integrated quality assurance. Screening is a complex process that relies on the optimal configuration and functioning of a number of components, including determination of the target population, formulation of the screening protocol, invitation coverage, efforts to maintain high attendance rates, validity of the screening test, and the follow-up and management of screen positives. Monitoring of performance indicators and, especially, process audits are needed for identification and rectification of any barriers to effectiveness in the screening chain. In addition, periodic evaluation of the effects of the screening activity on incidence and mortality endpoints is needed, as the risk factors for the population may change between areas and with time.

The aim of this study was to evaluate the performance and age-specific effectiveness of cervical cancer screening, by focusing on audit studies of the Finnish cervical cancer screening programme within case–control designs with information on the outcome of screening. The study also developed further quality assurance protocols for integration into the programme. Good quality of process and outcome registration is required for reliable quality assurance activities. The coverage and accuracy of data on screen-detected lesions and cancers in the screening register were evaluated via individual linkage to two other health care registers. Precancerous and cancerous lesions arising in the screened population were used in an audit of cytology wherein screening-test validity in programme service laboratories was evaluated with a focus on sensitivity failures in the form of false negative screening results. Cytodiagnosis sensitivity, specificity, and reproducibility were evaluated in a review phase involving both case and control smears. Outcome measures were also related to cross-sectional performance indicators, and variations between laboratories were explored.

The mode of detection and screening history was determined for every cancer case diagnosed in Finland in 2000–2009. A separate audit of screening histories was performed for deaths from cervical cancer in the same period. Population-based controls were used in estimation of the agespecific effectiveness of the organised programme in a case–control design. We were also able to perform age-specific self-selection bias corrections of the effectiveness estimates, which should ensure more informative results.

The quality of the screening and cancer registers is such that reliable monitoring and also individual case audits are possible. We found that some analytical

failures as measured in terms of false-negative rates of case smears do occur in the programme but that their impact on cancer incidence is small. However, the reproducibility of the cytodiagnosis and variations in the specificity of the screening laboratories should be addressed by means of cytology audits and feedback to the screeners.

A large proportion of the cervical cancers and most deaths from cervical cancer occur at ages above the currently recommended invitational ages. Only a very small proportion of the burden arises before first invitation. Non-attenders contribute significantly to incidence and mortality, and a smaller proportion of cases can be attributed to screening failures. Management of screening positives appears to be excellent. The effectiveness of screening, as measured by the reduction in the risk of cervical cancer and death from cervical cancer associated with participation in organised screening, was strongly dependent on age. Screening at ages below 40 and, especially, below 30 was associated with a clearly smaller risk reduction than screening at 40 and above. Also the duration of the protective effect was age-dependent. These findings refer specifically to participation in the organised screening programme against a backdrop of considerable opportunistic screening activity.

In light of these audit studies, one can see that most cancers and cancer deaths currently occur because of a lack of screening, either among non-attenders of the screening programme or women outside the screening ages, and not because of low quality of the screening test or management process. Variability of performance and, especially, of the laboratories' specificity still warrants regular feedback and harmonisation, so that the adverse effects caused by false positives are kept at a minimum. Monitoring and audits of the screening programme are clearly important for the programme's development and further optimisation. Also changing circumstances such as reorganisation of screening and the associated health service providers, and dynamic risk factors in the population require constant vigilance and the ability to detect and respond to developing threats to the effectiveness of the programme.

2 FINNISH SUMMARY

Kohdunkaulasyövän seulonnalla voidaan ehkäistä yli 80% syövistä ja syöpään liittyvistä kuolemista. Tämän tasoinen vaikutus vaatii kuitenkin toteutuakseen hyvin organisoidun seulontaohjelman jonka yhtenä olennaisena osana on riittävän systemaattinen laadunvarmistus. Tehokas seulonta on monen osatekijän summa. Seulontaohjelman tehoon vaikuttavat kohdeväestön määrittely, seulontakäytännöt, kutsujen peittävyys, osallistumisaste, seulontatestin osuvuus ja seulontapositiivisten jatkotutkimusten ja hoitojen laatu. Prosessia kuvaavien määreiden seuraaminen ja etenkin päätetapahtumiin perustuva auditointi mahdollistavat vaikuttavuutta rajoittavien tekijöiden yksilöimisen ja korjaamisen. Myös vaikuttavuusasteen arviointi ilmaantuvuus- ja kuolleisuusmittareilla esimerkiksi eri ikäryhmissä on tarpeen sillä populaatioryhmien riskitekijät voivat muuttua ajan myötä.

Tämän väitöskirjatyön tavoitteena oli tutkia kohdunkaulan syöpää ehkäisevän seulonnan diagnostinen laatu ja ikäspesifi vaikuttavuus. Tutkimus paneutui Suomen seulontaohjelmassa tapaus-verrokki –asetelmissa tehtyihin auditointitutkimuksiin, joissa seulonnan päätetapahtumista saatavilla oleva informaatio otettiin huomioon. Tutkimus myös kehitti seulontaohjelmaan pysyvästi liitettäviä laadunvarmistustoimintoja päätetapahtumien auditointiin perustuen. Rekisteritiedon laatu on olennainen edellytys luotettavalle laadunvarmistustoiminnalle. Seulontarekisterin seulontalöydöstiedon kattavuutta ja oikeellisuutta arvioitiin yhdistämällä tietoja kolmesta eri terveydenhuoltoon liittyvästä rekisteristä. Sytologisen seulontatestin auditointiin käytettiin esiaste- ja syöpätietoja. Testin osuvuutta arvioitiin erityisesti väärien negatiivisten osalta. Uudelleenluentavaiheessa arvioitiin myös seulontatestin herkkyyttä, tarkkuutta ja toistettavuutta. Seulonnan jälkeisten päätetapahtumien suhdetta seulontaprosessia kuvaaviin parametreihin arvioitiin laboratoriokohtaisesti.

Suomen 2000-luvun syöpätapahtumien diagnoositapa ja seulontahistoria määritettiin yhdistämällä syöpärekisterin kattavat syöpätiedot seulontarekisterin tietoihin. Väestöpohjaisten verrokkien avulla arvioitiin seulontatestin ikäryhmäkohtaista vaikuttavuutta ilmaantuvuuteen ja kuolleisuuteen. Ehkä ensimmäistä kertaa pystyimme myös määrittämään ikäryhmäkohtaisen osallistumisharhan jonka avulla korjattujen vaikuttavuusestimaattien absoluuttinen informaatioarvo todennäköisesti parani huomattavasti.

Seulonta- ja syöpärekisteritiedon laatu on korkeaa tasoa ja mahdollistaa näin ollen luotettavan laadunvarmistustoiminnan ja tapausten auditoimisen. Seulontaohjelmassa tapahtuu jonkin verran analyysivirheitä mutta näiden vaikutus syövän ilmaantuvuuteen on pieni. Sytologisen diagnoosin toistettavuus ja laboratoriokoh-

taisen testitarkkuuden vaihtelut kuitenkin vaativat prosessin tarkkailua, palautteen antamista ja toiminnan kehittämistä myös jatkossa.

Suuri osa kohdunkaulan syövistä ja suurin osa kuolemista ajoittuvat diagnoosiltaan seulontaohjelman viimeisen kutsun jälkeiseen aikaan naisen elämässä. Tämä kuvastaa osaltaan seulontaikäryhmiin kohdistuvan seulonnan vaikuttavuutta koska seulomattoman väestön kohdunkaulan syövästä aiheutuva taakka ajoittuu paljon nuorempaan väestöön. Toisaalta vain hyvin pieni osa syövistä diagnosoidaan ennen ensimmäistä seulontakutsua. Seulontaan osallistumattomat naiset muodostavat toiseksi suurimman joukon ja pienempi osa kohdunkaulan syövistä ja kuolemista kohdistuu seulontaan osallistuneiden joukkoon. Positiivisten seulontalöydösten hoidon ja seurannan laatu vaikuttaa erinomaiselta. Seulontaan osallistumisen ja päätetapahtumien riskin välinen yhteys oli ikäriippuvainen, siten että alle 40 vuoden, ja etenkin alle 30 vuoden iässä tapahtuvalla ohjelmaseulonnalla arvioitiin olevan selvästi pienempi vaikutus kuin 40:n ja sitä vanhempien seulonnalla. Myös seulonnan riskiä pienentävän vaikutuksen kesto kasvoi iän kasvaessa. Tutkimuksen tulokset kuvaavat nimenomaan organisoituun seulontaohjelmaan osallistumisen vaikutuksia. Ohjelman lisäksi ja organisatorisesti siitä riippumatta väestö on myös merkittävän opportunistisen seulontatoiminnan kohteena.

Tutkimustulosten valossa voidaan todeta että suurin osa kohdunkaulan syövistä ja niihin liittyvistä kuolemista johtuvat seulonnan puutteesta, jolloin nainen ei ollut osallistunut organisoituun seulontaan kutsun saatuaan tai oli iältään seulontaikäryhmien ulkopuolella, eikä niinkään seulonnan laadun puutteista. Laboratoriotoiminnan ja etenkin seulontatestin tarkkuuden vaihtelut kuitenkin vaativat palautejärjestelmää jotta toiminta voitaisiin yhdenmukaistaa ja vääristä positiivisista testivastauksista johtuvat haittavaikutukset minimoida. Seulontaohjelman vaikuttavuuden ylläpitäminen ja edelleen kehittäminen vaatii laadunvarmistustoimintaa ja syöpätapausten auditointitutkimuksia myös jatkossa. Myös olosuhteiden muutokset, kuten terveyspalveluketjujen organisaatiomuutokset ja riskitekijöiden muutokset väestössä, vaativat jatkuvaa valmiutta havaita ja puuttua seulonnan vaikuttavuutta uhkaaviin tekijöihin.

3 LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications, which are referred to in the text by their Roman numerals.

I Lönnberg S, Leinonen M, Malila N, Anttila A. Validation of histological diagnoses in a national cervical screening register. Acta Oncol 2012;51:37-44.

II Lönnberg S, Anttila A, Kotaniemi-Talonen L, Kujari H, Melkko H, Granroth G, Vornanen M, Pietiläinen T, Sankila A, Arola J, Luostarinen T, Nieminen P. Low proportion of false-negative smears in the Finnish program for cervical cancer screening. Cancer Epidemiol Biomarkers Prev 2010;19:381-7.

III Lönnberg S, Nieminen P, Kotaniemi-Talonen L, Kujari H, Melkko H, Granroth G, Vornanen M, Pietiläinen T, Arola J, Tarkkanen J, Luostarinen T, Anttila A. Large performance variation does not affect outcome in the Finnish cervical cancer screening programme. Cytopathol 2012;23:172-80.

IV Lönnberg S, Ahti Anttila, Luostarinen T, Nieminen P. Age-specific effectiveness of the Finnish cervical cancer screening programme. Cancer Epidemiol Biomarkers Prev 2012;21:1354-61.

V Lönnberg S, Nieminen P, Luostarinen T, Anttila A. Mortality audit of the Finnish cervical cancer screening programme. Int J Cancer 2012. doi: 10.1002/ijc.27844.

4 ABBREVIATIONS

AGC-NOS	atypical glandular cells – not otherwise specified
AIS	adenocarcinoma *in situ*
ASC-H	atypical squamous cells – high-grade cannot be ruled out
ASC-US	atypical squamous cells – undetermined significance
CIN1-3	cervical intraepithelial neoplasia, grade 1–3
CIN1-3+	CIN1-3 or worse
CIN3/AIS	CIN3 or adenocarcinoma *in situ*
ENCR	European Network of Cancer Registries
FCR	Finnish Cancer Registry (or Register)
HDR	Care Registers for Social Welfare and Health Care (formerly known as the Hospital Discharge Register)
HPV	human papillomavirus
hrHPV	high-risk HPV
HSIL	high-grade squamous intraepithelial lesion
ICC	invasive cervical cancer
ICD-10	International Classification of Diseases – 10th edition
ICD-O-3	International Classification of Diseases for Oncology – 3rd edition
LSIL	low-grade squamous intraepithelial lesion
MSR	Mass Screening Registry (or Register)
NHS	National Health Service (British health service provider)
NOS	not otherwise specified
NPV	negative predictive value
Pap I–V	Papanicolaou class I–V
Pap smear	cytological smear stained with the Papanicolaou method
PPV	positive predictive value
SCC	squamous cell carcinoma (of the cervix)
THL	Terveyden ja Hyvinvoinnin Laitos (Finland's National Institute of Health and Welfare)
WHO	World Health Organization

5 INTRODUCTION

Cervical cancer screening via sampling of cells directly from the cervix, endocervix, and vagina has over the past half-century proved to be a very effective way of preventing the development of invasive cervical disease. The detection of abnormal cells in cervical scrapings or pap smears and subsequent diagnostic confirmation by colposcopy allow the excision of lesions at risk of becoming malignant. Cervical cancer has several properties that make it particularly suitable for screening. It usually develops via premalignant lesions in squamous or glandular epithelial cells visible in the transformation zone and the junction between these two different epithelia. The transformation zone and junction is visible at younger ages at the ectocervix surface. At older ages, because of the metaplastic process, the junction often withdraws into the endocervical canal and may not be visible, even if still accessible for sampling. Hence, direct sampling of the potentially abnormal exfoliated cells can usually be performed without invasive procedures. Crucially, cervical cancer develops from a long-lasting and treatable premalignant phase that is suitable for detection by cytology and related histology.

Despite these advantages, the effectiveness of screening programmes has been widely variable. In some settings, large reductions in both incidence and mortality of cervical cancer have been achieved, but in many cases impacts have been disappointingly low or even absent (IARC 2005, Anttila et al. 2009, Arbyn et al. 2009b). In order for the beneficial effects of screening to materialise, the entire chain from screening coverage to quality of treatment and follow-up has to be carefully controlled and optimised (EC 2003, IARC 2005, Arbyn et al. 2008a). Comprehensive quality assurance for the screening process is possible only in the context of the organised and population-based screening programme. The quality of an individually tailored or opportunistic screening service can be excellent but is usually not consistently so and in any case remains very difficult to monitor, evaluate, and hence improve.

Comprehensive quality assurance for a screening programme can be viewed as comprising both monitoring and evaluation. Monitoring includes the tracking of cross-sectional performance indicators – for instance, as specified in the European guidelines – but also outcome-based longitudinal indicators such as interval cancer rates are needed (Arbyn et al. 2008a). The derivation of outcome indicators requires the linkage of screening and cancer registers and the construction of screening history for all cancers – in other words, an audit of cervical cancer cases. A complete audit also entails review of both cytological screening tests and any histological slides, preferably with controls for reduced subjective evaluation bias. However, there are

crucial questions related to the optimisation of a screening programme that cannot be addressed with information derived from monitoring alone (for example the appropriate screening ages and intervals in a particular population); hence, in addition to continuous monitoring of a programme, it is necessary that there also be periodic evaluation studies performed with a cohort–control or case–control design.

The ideal, then, is an organised population-based screening programme with quality assurance at every step in the screening chain; regular audits of all incident cancer cases; and evaluation of efficacy, effectiveness, benefit/harm ratio, and cost (Sasieni and Cuzick 2001, Arbyn et al. 2008a). Such a programme will continuously produce information about possible barriers to optimal impact and allow dynamic responses to changes in the quality of component functions or in the risk profile of the targeted population (Andrae and Smith 1999).

The objective of this work was to develop and pilot components of quality assurance for integration into the cervical cancer screening programme and in the process to gain knowledge of the barriers to effectiveness and areas showing potential for improvement.

6 REVIEW OF THE LITERATURE

6.1 CERVICAL CANCER AND INTRAEPITHELIAL NEOPLASIA

The uterine cervix is the part of the uterus that projects into the upper part of the vagina. Two types of epithelia converge at the cervix. Stratified squamous epithelial cells line the inside of the vagina and most of the ectocervical surface, and glandular epithelial cells line the endocervical canal and also mucus-secreting glandular crypts within the squamous epithelium. At the onset of puberty, a considerable area of the ectocervix is covered by columnar epithelium, but gradually the squamocolumnar junction starts to migrate from the periphery of the ectocervix toward the endocervical canal through squamous metaplasia. This is a slow replacement of the columnar epithelium by squamous cells. The area where squamous metaplasia takes place is called the transformation zone. The cells of the transformation zone are especially susceptible to HPV-induced neoplastic transformation, and this is where most squamous-cell carcinomas of the cervix develop (IARC 2005, IARC 2007). Similarly, columnar cells near the transformation zone and in the endocervical canal are believed to give rise to adenocarcinomas of the cervix.

Persistent infection by certain high-risk types of HPV is necessary for the development of both squamous and adenomatous cervical cancers (zur Hausen 1976, Bosch et al. 1995, Munoz et al. 2003). Of over 130 known types of HPV, about 40 can infect the genital mucosa, and 18 of the latter are termed high-risk types because of their ability to immortalise squamous epithelial cells (de Villiers et al. 2004).

Most sexually active women will contract a genital HPV infection at some point in life, and the prevalence is especially high at young ages, a few years after sexual debut (Rodriguez et al. 2007). In a cohort of university students in Finland, the prevalence of HPV in the lower genital tract was 33% (Auvinen et al. 2005). However, the majority of HPV infections are cleared by host defences within 6–12 months without complications, even when the causative agent is of an hrHPV type (Berkhof et al. 2005). When infection persists, there is a risk of neoplastic transformation and the development of precancerous lesions and cervical cancer.

Neoplastic cervical lesions of squamous-cell origin can be classified by the thickness of atypical, abnormal cells within the epithelium and by their cellular characteristics into cervical intraepithelial neoplasia (CIN) of grades 1–3. Lesions classified as CIN1 can be caused by both low- and high-risk types of HPV, and the risk of progression to cervical cancer is low. CIN2 and CIN3 are predominantly

associated with high-risk types, and the risk of progression is higher. An early Finnish study estimated that 28–39% of severe dysplasia and carcinoma *in situ* cases combined (equivalent to CIN3) will progress to invasive cancer (Hakama and Räsänen-Virtanen 1976). Progressive proportions observed in a study of untreated CIN3 cases in New Zealand with 30 years of follow-up (McCredie et al. 2008) and in a study using data from Canada's British Columbia screening programme (van Oortmarssen and Habbema 1991) fall within the same range (at 31% and 38%, respectively), though these proportions may be underestimates of true undisturbed progressive potential and show strong variation with age. The latter study also provided evidence of higher regression rates at younger ages.

The development of cervical cancer from hrHPV infection through stages of precancerous lesions is usually a slow process. A case–control study nested in a cohort of women screened in the Swedish screening programme estimated the mean incubation period from first confirmed HPV 16 infection to detection of carcinoma *in situ* at the cervix uteri to be 17–19 years, depending on the viral load (Ylitalo et al. 2000). In a modelling study based on data from the British Columbia screening programme, the average duration of the dysplasia and carcinoma *in situ* stages combined was estimated at 11.8 years (van Oortmarssen and Habbema 1991).

Squamous-cell carcinomas make up the bulk of cervical cancers, in the range of 80–90% in an unscreened population. The second most common group, the adenocarcinomas, originate in the columnar epithelial cells of the cervix. The natural history of this group is not very well charted, but an adenocarcinoma *in situ* (AIS) stage is commonly recognised. Some other histopathological types of carcinomas and a small number of stromal malignancies (mainly sarcomas) are also counted among cervical cancers.

6.2 SCREENING FOR CERVICAL CANCER

Screening is the use of methods of detecting unrecognised health risks or diseases in order to permit timely intervention (IARC 2005). The development of the first specifications for the requirements of screening was commissioned by the WHO and published in 1968 (Wilson and Jungner 1968). These principles are designed to ensure continuous and universal access to a screening service with the necessary biological and organisational conditions in place for a positive health impact at acceptable cost. More recently, the psychological and physical harm of screening, especially that related to false-positive tests, over-diagnosis, and over-treatment, has been brought into focus (Arbyn et al. 2008c, Hellsten et al. 2008).

The primary aim of cervical screening is to reduce incidence and mortality of cervical cancer. A secondary aim can be to improve cure rates via screen-detection of non-symptomatic cancers. There are three major types of screening tests suitable for achieving these aims. The predominant screening modality is based on the microscopy of exfoliated cells from the cervical epithelium. This technique was proposed by the Greek pathologist Papanicolaou (Papanicolaou and Traut 1941) and has been used to great effect in cervical screening. Modifications of this technique include automation-assisted analysis and liquid-based cytology (Nieminen et al. 2003, Arbyn et al. 2008b). Cytology-based screening tests are an attempt to identify abnormal cells that indicate the presence of a precancerous lesion. After diagnostic confirmation by colposcopy and biopsies, lesions can be excised, whereby the development of invasive disease is prevented. The second modality is based on detection of the presence of high-risk HPV. As persistent HPV infections precede the development of neoplastic cellular changes, the disease process may be detectable with an HPV test at even earlier stages than with cytology. The HPV tests in most common use are based on the detection of viral DNA from cellular material. Also RNA and viral protein assays have been developed, but these have not yet been clinically validated for cancer screening purposes. The third modality is visual inspection of the cervix, usually aided by acetic-acid treatment of the epithelium. This method is used in low-resource settings (Sankaranarayanan et al. 2007).

6.3 THE AUDIT OF CERVICAL CANCER CASES

6.3.1 DEFINITION

Cervical cancer screening is a complex procedure the outcome of which depends on seamless collaboration of several functions. The performance of the process can indirectly be observed in trends of cervical cancer incidence and mortality, but if one is to develop and optimise the programme, detailed information about the performance of component functions is needed. As with any large-scale process, an effective way of gaining useful insights into the performance is to examine the failures of the process (Cuzick 2008). The failures of cervical screening are women who develop (potentially fatal) cervical cancer.

Screening histories of cervical cancer cases have long been of interest to screening organisers and screening professionals, and there are examples of published case series from the 1970s and 1980s (Rylander 1976, Grundsell et al. 1979, Dunn and Schweitzer 1981, Brown and Barker 1982, Yajima et al. 1982). Proposals for

more systematic audits including smear reviews were probably first made in the UK in the 1980s (Chamberlain 1984), resulting in pilot audit studies in the 1990s (Slater et al. 1994, Sasieni et al. 1996) and in formalised audit specifications being published in 2006 by the NHS (NHS Cervical Screening Programme 2006), as well as in the European guidelines for quality assurance in cervical cancer screening, in 2008 (Arbyn et al. 2008a). A Swedish study in 2008 was hailed as the first comprehensive national audit since publication of the European quality assurance guidelines (Andrae et al. 2008). According to specifications in these publications, a comprehensive audit starts with a screening history review of all invasive cervical cancer cases arising in the population covered by the screening programme within a specific span of time. Any (negative) smears preceding the diagnosis should be reviewed for failures of analysis (cytology audit). Any histological slides preceding the diagnosis could likewise be reviewed (histology audit). Controls can be used to alleviate the tendency to overcall audit smears that are known to precede a cancer diagnosis (Renshaw et al. 2004, Coleman and Poznansky 2006) but also for uncovering information on the specificity of the screening test and histological evaluation (Arbyn et al. 2008a). The controls can be population-based controls, with sampling from women at risk of cervical cancer (for example, matched for age and place of residence), in which case the same controls can be used for a case–control analysis of the association of screening history with the risk of cervical cancer (Arbyn et al. 2009a). Alternatively, direct archive specimen–control sampling can be used for the cytology and histology reviews (Repse-Fokter et al. 2012).

6.3.2 AUDIT OF SCREENING HISTORY

For reliable monitoring and audit of a cervical screening programme, it is essential to have complete registration of all cervical cancer cases and all preventive screening tests in the population. In addition, a reliable means of linking registers (e.g., a personal identifier) is needed. Screening history can be categorised with various levels of detail, depending upon available information and anticipated barriers to effectiveness in each particular setting. However, screen-detected cancers should be categorised according to the previous screening history, as the screening event leading to detection does not afford any protection. Mode of detection (by screen or by symptoms) is an important parameter independently associated with both stage and survival (Herbert et al. 2009b, Andrae et al. 2012), and it has been recommended for inclusion in routine cancer registration (ENCR 2001). It is not useful in the analysis of preventive screening history or screening failure audit, however. Table 1 illustrates a hierarchically organised synthesis of the types of failures that are of interest with respect to screening-history audits (Hakama et al. 1985, Miller 1995, Zapka et al.

2003, Bagnall et al. 2006). The example details a screening history categorisation scheme that classifies each cancer case with respect to a defined period of time before diagnosis (for instance, one recommended screening interval). Every cancer case can be placed into one of the three classes in the first column in the table. The classes are mutually exclusive, and the classification proceeds in order such that only those cases not falling into the first class are considered for the next one. Those not invited include women who, because of age or incomplete invitational coverage of the target ages, were not extended an invitation in the specified time window before diagnosis. Those invited can then be classified according to whether they participated and a preventive screening test was performed. Participating women can further be classified by screening test result, confirmation result, appropriateness of follow-up, and treatment. If a large, or increasing, proportion of cases fall into a specific category of screening history, a targeted evaluation should be triggered. For example, if most cancers were to appear outside the targeted screening ages, investigation into the possibility of extending the target age range might be useful.

An approach commonly used for categorising process failure is the division of cancer cases by screening history into failures of coverage (by invitation or participation), failures of detection (negative screening tests), and failures of management (failures of follow-up and treatment) (Spence et al. 2007). False-negative histopathology could be taken to be a failure of detection or failure of management; usually it is considered to involve the latter. However, a more detailed specification will allow for more effective feedback to the screening service providers, which is a necessary component of an audit aimed at improvement in the effectiveness of the screening programme.

Most audits include all invasive cancer cases in a given population diagnosed during a defined period of time. However, some authors do not consider microinvasive carcinomas to be failures of screening, because of their extremely favourable prognosis (IARC 2005), and focus instead on frankly invasive cancers (stages IB and above). Also, audits including only fatal cases of cervical cancer have been published (Wilson and Johnson 1992, Slater et al. 1994, Mitchell et al. 1996) but not recently. Furthermore, some authors consider only SCCs to be failures of a screening programme and hence suitable as audit cases, as the effectiveness of cytological screening has been unequivocally established for that particular histopathological subspecies (IARC 1986) and smaller or non-existent protective effects have been observed for adenocarcinomas (Herrero et al. 1992, Makino et al. 1995, Mitchell et al. 1995, Mitchell et al. 2003, Zappa et al. 2004). As there is a differential association of screening and cervical cancer by stage and histopathology, it is advisable to present audit and evaluation results for frankly invasive cancers and SCCs separately, in addition to results for all cervical cancers.

Table 1: Screening history categorization for a cancer audit (failure analysis)

- Not invited
  - Not targeted by programme
  - Invitation failure
- Participation failure
- Interval cancer
  - Negative screening test
    - Analytical failure (false negative)
    - Sampling failure
    - Incompatibility of test modality and natural history*
  - Borderline screening test
    - Analytical failure (false negative)
    - Failure of follow-up
  - Inadequate sample
    - Failure of follow-up
  - Referral
    - Referral compliance or organisation failure
    - Negative histology
      - Analytical failure (false negative)
      - Sampling failure
    - Positive histology
      - Ineffective treatment of CIN
      - Failure of follow-up

* Inherent test sensitivity is too low or interval is too long or both.

Table 2 lists research papers that describe the screening history of case series or case and control subjects where the purpose has specifically been audit research – i.e., to elucidate the process of care failure that allowed cervical cancers to develop. These papers were selected for comparability of the proportion of cases that were interval cancers, defined here as cases arising after a screening test with any outcome but not leading to the detection of cancer, in a specified interval before diagnosis. Also, the papers listed consider all cervical cancers or all squamous-cell cervical cancers diagnosed in a specified population during a specified time and use screening or cytopathology registers for the construction of screening history.

The mode of detection with respect to screening can be challenging to assign correctly, and various approaches have been used, in different settings. In the Netherlands, information on the mode of detection is available in the automated national pathology archive (PALGA), and diagnostic and preventive screening tests can be correctly distinguished (Bos et al. 2006, van der Aa et al. 2008, de Bie et al. 2011). In the audit in Lambeth and Southwark, London, it was also possible to distinguish screen-detected cancers by clinical criteria, with the definition being those diagnosed as a result of investigation of an abnormal cytology test (Herbert et al. 2010). This information is now routinely collected for each woman with cancer in the London quality assurance database. When not directly recorded in pathology

or cancer registers on the basis of referral information, the identification of smears leading to screen-detection of cancer, as opposed to preventive smears, may be difficult. Some studies try to alleviate the problem by considering only frankly invasive cancers (stages IB+), following the rationale that microinvasive cancers are usually screen-detected (Anderson et al. 1992). Other studies exclude the screening history immediately before diagnosis (for example, six months before), considering smears taken during that time to be part of the diagnostic process (Kenter et al. 1996, Andrae et al. 2008, Ingemann-Hansen et al. 2008, Herbert et al. 2009a, Kirschner et al. 2011). A third option is to regard all cases of cancer diagnosed within a certain period after referral for colposcopy based on a positive screening test to be screen-detected, modifying the time interval for screening history classification accordingly. All of these approaches yield approximations only. The first approach is compromised by the inevitable occurrence of frankly invasive screen-detected cancers, the second by the fact that screening events occurring close to diagnosis but not leading to detection (potentially false-negative screens) are missed when this period is excluded. These are important indicators of the process of care failure. Finally, the third approach can be unreliable if there is non-compliance with referral.

The first study in this overview is a Danish report describing the screening histories of 376 women with cervical cancer diagnosed between 1979 and 1983 (Kristensen et al. 1991). In the three years before diagnosis, 202 of the women, or 54% of all cases, had been screened. However, it is not clear whether the histories of all 202 women included preventive screens – i.e., screening tests in addition to those leading to the detection of cancer. A later, and more detailed, Danish audit reported screening histories of 286 ICCs in Aarhus County 10 years after the implementation of an invitational screening programme for women aged 23–59 (Ingemann-Hansen et al. 2008). The screening history was based on tests five to 47 months prior to diagnosis, a suitable time window for preventive smears. Across all ages, 23% of the cases never had screening, another 38% were not screened in the last interval, 20% involved interval cancers with a preceding negative test, 5% manifested inadequate follow-up after an abnormal smear, and 7% involved development of cancer despite adequate management of an abnormal smear. The authors conclude that improving participation in the programme should be given high priority.

Anderson and colleagues (1992) presented a failure analysis audit from British Columbia with 437 ICC cases of stages IB and above diagnosed in 1985–1988. All cases in the province were referred to one of two clinics, in Vancouver and Victoria, and all screening tests were analysed at one central screening laboratory and recorded in a central cytology register. Hence, the completeness of data on both cancer cases and screening tests can be expected to be good. However, no description of the method of linkage was given in the report. There were 170 cases, 39% of the total, with no cytology examinations recorded and a further 10% with more

than five years since last cytology. The screening history was defined as screening tests before the presentation of invasive disease, which is somewhat vague in terms of the handling of screendetection. Probably the only precaution against including diagnostic screening tests of screen-detected cancers was the exclusion of micro-invasive carcinomas.

Similar results were later reported from the Canadian province of Alberta, where the screening history and failure category of cases diagnosed in 1990–1991 were reported (Stuart et al. 1997). Out of 246 cases, 30% had never been screened; a further 15% had been screened more than three years before diagnosis; and in 17% of the cases, the screening history could not be ascertained. Another 10 years later, a third report from Canada reported screening histories of cases diagnosed in 2001–2002 in Toronto (Spayne et al. 2008). Here, 31% of the 225 ICC cases analysed involved a screening test within 6–48 months before diagnosis. The three Canadian studies show a downward trend in the proportion of cancers with screening tests within 3–5 years before diagnosis: 51% in British Columbia in the 1980s, 43% in Alberta in the 1990s, and 31% in Ontario in the 2000s. It is possible that this trend reflects improvements in the quality of cytology and management services over time.

There are a number of audit studies from the UK. An early-screening failure analysis included 36 cervical cancer deaths in Rotherham (Slater et al. 1994). Screening history was determined for a period of eight years before diagnosis, and 53% of the women had been screened in this time window. One of the early pilot audit studies from the UK analysed screening histories of 348 invasive cancers from 24 self-selected health districts (Sasieni et al. 1996). Tests from six months before diagnosis were excluded. In this material, 53% of cancers were interval cancers with a screening test within 66 months of diagnosis. Of all fully invasive cancers, 48% were interval cancers, and the corresponding proportion for all microinvasive carcinomas was 69%. A later study from the UK investigated the screening histories of cervical cancer cases in a 12-year period during introduction of organised screening in Southampton and south-west Hampshire (Herbert et al. 2009a). There were 382 women with incident cervical cancer in 1986–1995. Cytology tests performed within six months of diagnosis were considered diagnostic and excluded from the screening history. Interval cancers, or cases with screening tests within 5.5 years of diagnosis, accounted for 45%, and 22% of these had a negative last screen. The study also presented an interesting figure showing proportion of interval cancers as a function of time since programme implementation. The proportion climbs initially as coverage is expanded, peaks six years later, and then starts to fall alongside overall incidence. The proportion of cases in non-participants declines initially and then seems to reach a plateau. A year later, that work's first author published a similar study of CIN2+ cases diagnosed in two south London boroughs (Herbert

et al. 2010). In addition to 3,027 precancer cases, 133 invasive cancer cases were diagnosed in 1999–2007. Of the cancers, for 53% there had been a screening test 0.5–5.5 years previously. A large proportion of cancers were screen-detected in this audit (49%); two thirds of these were microinvasive.

In the US, screening audits have been challenging because of a lack of personal identifiers and comprehensive screening registers. Nevertheless, some screening failure analysis reports have been published. One study audited 664 cervical cancer cases diagnosed in Connecticut from March 1985 through February 1990 (Janerich et al. 1995), by collecting screening history information from physicians, patients, and kin. Out of the 481 with screening history available, 48% had a screening test within five years prior to diagnosis, and 25% within three years. The Kaiser Permanente Medical Care Program of northern California provides a good setting for monitoring and evaluation of screening in the US although data are available only for health plan members. Between 1988 and 1994, 455 eligible ICC cases were diagnosed among long-term members, and their screening histories for 6–26 months before diagnosis were elicited (Sung et al. 2000). Non-participation was the largest category, with 53%, followed by 28% potential false negatives, 9% abnormal smears with correct management, and 4% with inadequate follow-up of cytological abnormalities. Participation was highest in the younger age groups. Interventions to increase participation were urged by the authors.

The screening programme in the Netherlands targeted women of ages 35–53 from 1998 and 30–60 since 1996. Of 401 ICC cases diagnosed in 1991–2008 in Nijmegen, 11% were diagnosed in women younger, and 22% older, than the target cohort (de Bie et al. 2011). Another 40% were not screened in the last five years, 21% had potentially false-negative tests, and the remaining 6% were screened as abnormal but follow-up or treatment had failed. Diagnostic smears were marked at registration and excluded from the study. Another Dutch audit analysed 2,074 cases of ICC, with 12% diagnosed before screening started and 20% at older ages than the target cohort and more than six years after last invitation (Bos et al. 2006). In this study only 19% had a preventive screening test in the last six years before diagnosis.

The Swedish audit used well-defined screening histories for 1,230 cases diagnosed in 1999–2001 and their age-matched controls (Andrae et al. 2008). National registers for cancer and screening episodes were linked with personal identifiers, ensuring high coverage and quality of data. Tests made within six months of diagnosis were considered diagnostic and excluded from the screening history. The proportion of cases with a negative screening test during the two screening rounds (six years) before diagnosis was 24%, and a further 11% had abnormal smears with or without biopsy, for at total of 36% with a history of screening. The presumed mode of detection was also reported, with 25% of cases being screen-detected. Another Swedish audit study, published in the same year, presented screening histories of

Table 2: Characteristics and results of reports on cervical cancer screening histories (failure analysis)

First author	Year	Country/region	Cases	Diagnosis date	No. of cases	SCC (%)	IA (%)	Cases with screen history	Follow-up period (y)	Screened	Previous screen neg	Comments†	
												Cytology review	Case–control analysis
Kristensen	1991	Denmark/Funen	ICC	1979–1983	376	82%	-	376	3	54%	31%	yes	no
Anderson	1992	Canada / British Columbia	ICC IB+	1985–1988	437	74%	0%	437	5	51%	26%	yes	no
Slater	1994	UK/Rotherham	ICC deaths	1989–1991	36	-	-	36	8	53%	39%	yes	no
Janerich	1995	USA/Connecticut	ICC	Mar. 1985–Feb. 1990	664	80%	-	481	5	48%	37%	yes	no
Kenter	1996	Netherlands/western	SCC	1980–1989	469	100%	-	306	3.5	27%	13%	yes	no
Sasieni	1996	UK / selected districts	ICC	1992	348	-	26%	348	5	53%	24%	no	yes
Stuart	1997	Canada/Alberta	ICC	1990–1991	246	79%	15%	197	3	43%	33%	yes	no
Sung	2000	USA / northern California	ICC	1988–1994	642	68%	-	455	3	47%	28%	no	no
Bos	2006	Netherlands	ICC	1994–1997	3,175	-	-	2074	6	19%	9%	no	no
Spayne	2007	Canada/Ontario	ICC	Apr. 2001–Mar. 2002	225	67%	-	225	4	31%	-	no	no
Andrae	2008	Sweden	ICC	1999–2001	1,230	75%	20%	1230	3.5–5.5*	36%	24%	no	yes
Ingemann-Hansen	2008	Denmark/Aarhus	ICC	1997–2002	286	81%	30%	286	3.5	40%	20%	no	no
Lindqvist	2008	Sweden/Malmö	ICC	1991–2000	187	83%	-	187	4–6*	39%	-	no	no
Yang	2008	Australia / New South Wales	ICC in ages 20–69	2000–2003	877	63%	-	877	4	33%	23%	no	yes
Herbert	2009	UK / Southampton and south-west Hampshire	ICC	1985–1996	382	76%	14%	382	5.5	45%	22%	no	no
Herbert	2010	UK / London boroughs	ICC	1999–2007	133	85%	35%	133	5	53%	20%	yes	no
de Bie	2011	Netherlands/Nijmegen	ICC	1991–2008	401	77%	7%	401	5	27%	21%	yes	no
Kirschner	2011	Denmark/Copenhagen	ICC	2008–2009	112	73%	39%	112	5.5	54%	-	yes	no
Repše-Fokter	2012	Slovenia	ICC	2006	162	-	26%	162	3	42%	30%	yes	no

* Recommended screening interval differs by age

† Studies with cytology review are also described in Table 3; studies with case–control analysis are also described in Table 4.

all 187 ICCs diagnosed in the city of Malmö between 1991 and 2000 (Lindqvist et al. 2008). There were 72 cancers among participants and, in addition, 16 women had declined further management after abnormal cytology, for, in all, 47% interval cancers. The screening results were not comprehensively reported, but 21% of all cases were classified as misread as normal in the smear review.

In New South Wales, 877 invasive cancers, in total, were diagnosed in 2000–2003 at ages of 20–69 (Yang et al. 2008). Each case was matched with three controls by month and year of birth. Goodquality registers of population, cancers, and screening tests were linked by means of probabilistic linking software using a number of non-unique personal data items. By screening history in the last four years, 66% of the cases were those of non-participants, compared with 13% of the controls. Pap tests of up to three months before diagnosis were excluded from the analysis. The authors recommend efforts to increase participation.

The practical conclusion of the failure analysis in nearly all of the studies reviewed was that increasing participation in screening offers the best potential for improving the effectiveness of the programme.

Studies including case–control analysis of the association between screening history and cervical cancer generally specified the derivation of screening exposure in greater detail. This involved mainly the precise and uniform definition of the period of screening history under observation and the exclusion of diagnostic smears.

It has been proposed that effective screening for cervical cancer leads to declining proportions of squamous-cell cancers in the screened populations, and this has been observed in the cervical cancer trends in many countries with efficient screening programmes (Finnish Cancer Registry 2011). A slight but non-significant trend of a decrease in SCC as a proportion of total cancers detected in regions and countries with presumably fairly well-established screening programmes was observed over time across the audit studies discussed (see Figure 1A).

The trends over time of interval cancers, according to the definition in Table 1, and interval cancers with a negative last smear reported in the studies listed in Table 2 are shown in Figure 1 (B–C). Even though declining trends are observed, neither measure shows statistically significant progress over time, possibly due to scattering, as the individual data points are derived from very different settings. In addition, the proportion of interval cancers depends not only on the effectiveness of the intervention itself but also on, for instance, the coverage of the at-risk population by the screening programme. On average, coverage has expanded over time, which may counteract the effects of improvements in screening intervention effectiveness on the proportion of interval cancers.

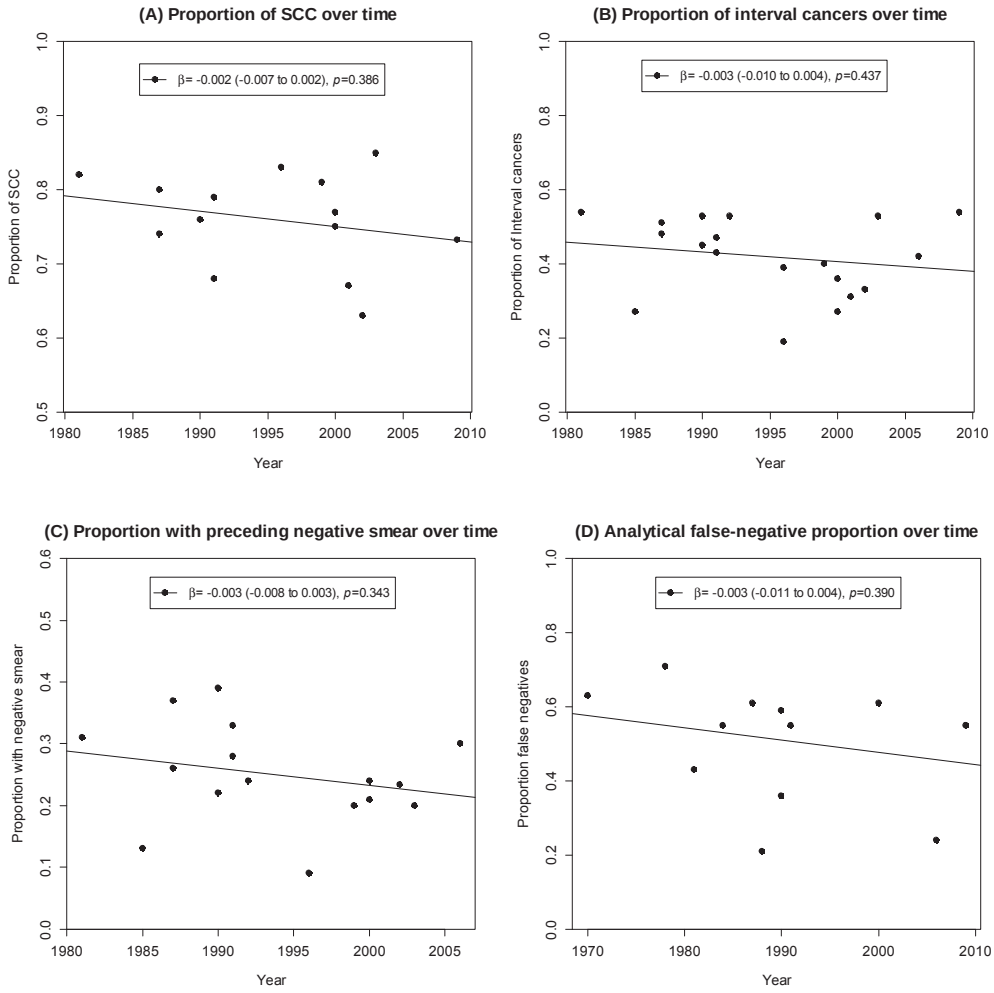

Figure 1. Proportions of SCC (A), interval cancers (B), and cases with a negative screening before diagnosis (C) out of all diagnosed cervical cancers, and the proportion of analytical false-negative smears out (D) of all reviewed negative smears preceding diagnosis, over average year of diagnosis in the study population. Linear regression coefficients, with the associated 95% CI and p value of the observed trends, are indicated.

6.3.3 AUDIT OF CYTOLOGY

The audit of cytology is concerned with those interval cancers that are preceded by a negative (or borderline) screening test (see Table 1). All negative screening tests in the preceding interval are potential false negatives. If the negative test result is confirmed in the review, the failure is due to either non-representative sampling or an incompatibility of screening test, test interval, and natural history of the specific case. It has been suggested that some cervical cancers may develop very

rapidly, such that one usually sufficient screening interval can encompass the whole development process from undetectable or absent disease to full-blown invasion (Austin and Zhao 2012).

Table 3 summarises cytological audit studies that are concerned with the sensitivity of the screening test, define potential false negatives as negative screening tests within a specified time window before diagnosis, and include ICCs or SCCs with or without microinvasive carcinomas in the base population. The false-negative proportion shows high variability between studies with review of negative (Pap I) smears (21–71%). Some improvement in the quality of cytology over time by this measure was observed, but the trend was not significant in linear regression (see Figure 1, pane D). The proportion of clearly positive (warranting referral for colposcopy) false negatives was less variable and ranged from 14% to 31% in the studies that did not include Pap II smears in the review. Usually, not all archive smears representing potential false-negative screening tests in a study can be located; for this reason, the denominator used for the false-negative proportion in the table is the number of negative smears reviewed and not the total number of cases in the audit.

One of the first studies that elegantly described the screening history and false-negative proportions of cervical cancer cases was that by Eva Rylander, with 177 cases diagnosed in 1968–1974 in Stockholm. Out of 69 negative smears four to five years prior to detection of cancer, 56 were retrieved for review. The review was done blinded with negative control samples from women without subsequent cancer diagnosis. Twenty-one of the smears reviewed were confirmed as negative, and 35 (63%) were judged atypical at presentation. This study also observed that interval cancers after negative smears were more frequent in younger women than among older women. The false-negative proportion was not presented by age, unfortunately.

An Australian report from Victoria audited 1,044 incident cases of cervical cancer of women aged under 70 (Mitchell et al. 1990). There were 156 negative cervical smears reported for the 36 months before the diagnosis of cancer, and 143 were obtained for review by a senior scientist and a cytopathologist. No controls were used. Seven smears (5%) were considered unsatisfactory for analysis, 21% had changes less than indicative of CIN, 31% showed changes indicative of CIN, and 43% were confirmed negative.

Kristensen et al. (1991) reported screening histories of 376 women from Funen, Denmark, who were diagnosed with cervical cancer in 1979–1983. Of these, 202 had smears taken within three years prior to diagnosis, with a total of 355 smears. No controls or blinding was used in the review. Of the 96 potentially false-negative smears, 57% were confirmed negative, 2% inadequate, and 18% atypical, and 23% were upgraded to CIN or cancer, for a false-negative proportion of 43%. A more recent study from Copenhagen described the results of a quality-control audit of 112 cancer cases diagnosed in 2008–2009 (Kirschner et al. 2011). There were 56 cases

with a regular screening history 5.5 years prior to the diagnosis of cancer, and 11 of the cases (20%) had smears reviewed as HSIL or worse. There was no mention of lower-grade atypical findings, which would presumably elevate the estimate if reported. No controls were used in this review.

In Canada, the screening programme in British Columbia was one of the earliest effectively organised programmes in the world (Fidler et al. 1968). A study of screening histories of 437 invasive cancer cases diagnosed in 1985–1988 reviewed all 116 negative smears in the screening files preceding diagnosis (Anderson et al. 1992). The review process was not described with respect to the number of reviewers, any controls, and blinding. False negatives, defined as missed abnormal cells, numbered 39 out of the 116 originally negative smears. There were 64 cases with negative smears within three years of diagnosis, and the authors stated that the false negatives were usually seen in this group, without reporting numbers. Another Canadian study reported screening histories for 246 invasive cancer cases diagnosed in Alberta in 1990–1991 (Stuart et al. 1997). In this study, several independent evaluations were used, with final cytology resolved by consensus of three pathologists. Controls and blinding were also used, to reduce observer bias. There were 170 smears with normal or benign (below ASC-US) results in the original analysis, and 93 (55%) of these were ASC-US or worse or were deemed unsuitable for interpretation at review. Review results requiring referral were recorded in 42 (25%) of the cases.

A review of negative cytology preceding frankly invasive cancers diagnosed in Northern Ireland in 1965–1989 found 139 negative smears from 103 patients up to 12 years before diagnosis (Robertson and Woodend 1993). The review was performed by the two authors independently, without controls. There were 92 smears with dyskaryotic cells (66%) and 67 out of 95 (71%) within five years of diagnosis.

One of the few audits concerned with cases of cervical cancer death is from the UK. It reviewed negative smears preceding the diagnosis of cervical cancers in Rotherham (Slater et al. 1994). There is no mention of blinding in the report. There were nine cases out of 36 with a true-negative last smear in the period examined (eight years prior to diagnosis) and five with a false-negative last smear (36% false-negative proportion). Three of these five were considered inadequate upon review, leaving two that were upgraded to positive samples (14% of the potential false negatives).

One large – if not the largest to date – published cytology audit used data from the cervical screening programme in England, with 6,113 cases of invasive carcinoma of the cervix diagnosed between April 2007 and March 2010 (Castañon et al. 2012). In total, 7,621 cytology samples taken within 10 years of diagnosis were reviewed, and 3,759 of these were potential false-negative screening tests. The review confirmed the negative result for 45% of these samples, 11% were deemed

inadequate for analysis, 18% were borderline, and 25% were clearly positive. There was no blinding of reviewers to the original results or the fact that the patient was subsequently diagnosed with cancer. This may have elevated the analytical false-negative rate observed. In fact, given that, of all studies reviewed, this one involved the longest time between sample and cancer diagnosis, the false-negative rate seems somewhat higher than expected. There were also 644 borderline samples in the review, of which 65% were deemed clearly positive upon review. Another British, welldescribed audit summary, from the Southampton area, presents review results for 73 negative slides taken up to 5.5 years before diagnosis (Herbert et al. 2009a). Of the potential false-negative slides, 11% were reviewed as inadequate, 17% as low-grade dyskaryosis, and 31% as high-grade dyskaryosis.

The only study in this overview from the US audited 664 cervical cancer cases diagnosed in Connecticut from March 1985 through February 1990 (Janerich et al. 1995). In all, 137 negative smears were obtained for review, and 29 (21%) of these were classified as 'misread as normal' on the basis of 'some evidence of dysplasia or premalignant abnormalities'. The review was performed without controls by two cytopathologists who were blinded to the screening history.

Kenter et al. (1996) reported audit results for 306 out of 469 patients diagnosed with SCC in 12 hospitals in the western part of the Netherlands in 1980–1989. All previous smears were reviewed with blinding to the original cytology, but no controls were used. There were 39 Pap I/II screening tests within 3.5 years of diagnosis, and 53% of these were considered a false negative, defined as a result requiring immediate referral. The remainder consisted of Pap II and inadequate smears upon review, and no smears were confirmed negative. In a later Dutch study, 401 invasive cervical cancers diagnosed between 1991 and 2008 at a hospital in Nijmegen were audited (de Bie et al. 2011). There were 85 patients with smears 'within negative limits' in the five years before diagnosis. No controls were used in the review, which confirmed negative results in 39% of cases, considered 24% unsatisfactory for analysis, deemed 18% ASC-US or LSIL, and found 19% to be HSIL or worse.

The Slovenian organised cervical screening programme was launched in 2003, and a cytological audit study of the 162 cervical cancer cases diagnosed in 2006 was published recently (Repše-Fokter et al. 2012). There were 34 normal smears by original cytology recorded in the three years prior to diagnosis, and eight of these were upgraded in a blinded review with controls, for a false-negative proportion of 24%. The upgraded diagnoses were not specified for this subgroup.

Table 3: Characteristics of studies reporting cytology audit results

First author	Year	Country/region	Cases	Diagnosis date	No. of cases	No. of negative slides reviewed	Original classification	Retrospective follow-up (years)	Controls in review	False negatives*	Clearly positive false negatives†
Rylander	1977	Sweden/Stockholm	ICC	1964–1974	171	56	Pap I	5	yes	63%	-
Mitchell	1990	Australia/Victoria	ICC under age of 70	1982–1986	1,044	136	negative	3	no	57%	31%
Kristensen	1991	Denmark/Funen	ICC	1979–1983	376	96	negative	3	no	43%	23%
Anderson	1992	Canada / British Columbia	ICC IB+	1985–1988	437	64	Pap I	3	no	61%	-
Robertson	1993	Northern Ireland	ICC IB+	1965–1989	103	95	negative	5	no	71%	-
Slater	1994	UK/Rotherham	ICC deaths	1989–1991	38	14	negative	8	no	36%	14%
Janerich	1995	USA/Connecticut	ICC	Mar. 1985–Feb. 199(	664	137	negative	5	no	21%	-
Kenter	1996	Netherlands/western	SCC	1980–1989	469	39	Pap I/II	3.5	no	100%	53%
Stuart	1997	Canada/Alberta	ICC	1990–1991	246	104	normal	3	yes	55%	25%
Herbert	2009	UK/Southampton and south-west Hampshire	ICC	1985–1996	382	73	normal	5.5	no	59%	31%
de Bie	2011	Netherlands/Nijmegen	ICC	1991–2008	421	85	normal	5	no	61%	19%
Kirschner	2011	Denmark/Copenhagen	ICC	2008–2009	112	58	negative	5.5	no	-	20%
Castañon	2012	England	ICC	Apr. 2008–Mar. 2010	6,113	3,759	normal	10	no	55%	25%
Repše-Fokter	2012	Slovenia	ICC	2006	162	34	normal	3	yes	24%	-

* Cytological abnormalities or inadequate sample

† Cytology requiring immediate referral for colposcopy.

Length of retrospective follow-up was analysed as an explanatory variable for the differences in false-negative proportions, but no correlation was found in this collection of audit studies. Neither did the use of controls appreciably explain the variability. The unweighted-average false-negative proportion reported in studies using controls in the review was 47%, as compared with 52% in the studies not using them. The corresponding averages weighted for number of smears reviewed were 52% and 54%. It is tempting to conclude that the differences seen in the false-negative rates really do reflect differences in the sensitivity of cytology between settings.

Many of the studies describing cytological audits of potentially false-negative smears were not reviewed here, because of attempts to keep the included studies comparable. The majority of the ineligible studies did not specify the interval between smear and diagnosis. An open-ended retrospective screening history provides more potential false-negative smears for review, but the exact length of time before diagnosis can be important for failure analysis. When the period examined before diagnosis becomes more than one or two screening intervals, the proportion of true negatives should be progressively higher and the significance of the results may diminish.

6.3.4 AUDIT OF HISTOPATHOLOGY

There are only a few studies that audit histopathology. While histopathology is often seen as a gold standard for screening-test validity studies, it can have validity problems of its own. The recent Danish audit study included review of negative histopathology slides and found two cases of false negativity in an audit case population of 112 (Kirschner et al. 2011). The number of potential false negatives reviewed was not reported. The large UK audit by Castañon et al. (2012) reviewed all histopathology specimens of the audited cancer cases, regardless of original diagnosis. Of 112 originally normal histological samples reviewed, 66% were confirmed normal in review, 13% were inadequate, and 16% were upgraded to intraepithelial neoplasia and 5% to invasive carcinoma. So with the strictest definition of a false negative, 34% of the histopathological diagnoses audited were false negative, as compared with the 55% for cytology false negatives in the same study. Reviews of histopathology should be included in the audit of cervical cancers, especially if screening history indicates that interval cancers occur after a negative colposcopy.

6.4 OUTCOME-BASED EVALUATION OF SCREENING PROGRAMMES

6.4.1 METHODOLOGY

The screening history of case series alone cannot answer questions as to the effectiveness of the programme, or loss of effectiveness due to a specific category of failure. The proportion of failure categories such as subsequent (within-interval or at-next-screen) cancers in the population will depend on the effectiveness of the screening service but also on the coverage of the programme. For this reason, it becomes necessary to include controls without cervical cancer in the analysis of screening histories when attempting to answer questions specifically about effectiveness and preventable percentages. Also, full case–cohort information would be appropriate, even though such analyses would require considerable resources for tracing all of the details in the screening history among non-cases; therefore, such designs are not considered here.

In case–control evaluations, it is important to pay close attention to the selection of cases and controls, the definition of exposure, and the handling of bias (Cronin et al. 1998). Specific concerns include inclusion of all case subjects within a source population, definition of eligibility criteria in a manner ensuring that the case and control subjects had equal access to screening during the exposure period, the distinction between symptomatic and screening smears, handling of screendetected cases of invasive cancer, and control of self-selection bias (Moss 1991, Cronin et al. 1998).

The case subjects should have a manifestation of disease that develops only after the preclinical stage of disease (Morrison 1982). In cervical cancer screening, this can be interpreted as meaning either fully invasive carcinoma of FIGO stage IB and above or, alternatively, any invasive carcinoma, which includes also stage IA. The argument against including microinvasive (IA) carcinomas in the endpoint definition is that they usually do not cause any symptoms and hence can be classified as falling under the preclinical stage of disease. On the other hand, if the aims of screening include the prevention of any invasive carcinoma of the cervix, all stages of invasive disease should be included in the evaluation, for a complete overview of effectiveness. One can use endpoints of later stages – for example, cancers that cause death. In evaluation of recently commenced screening programmes, cases occurring soon after initiation should be excluded, because full screening benefits would not be expected for several years (Moss 1991). This is especially important if the outcome is mortality.

Controls should be drawn from the source population of the cases. Control subjects are usually matched to the cases with respect to age, place of residence, and sometimes also socio-economic status. The matching is an attempt to equalise differences in factors that are associated with both exposure and risk of disease. Stratified analysis by age can be an alternative to matching.

If the aim is to evaluate the preventive effect of screening, all screening tests, both negative and positive, must be counted toward exposure. Tests done because of symptoms and true screening tests should also be distinguished. This can be problematic in programmes that are not based on invitation on a fixed date. In addition, tests leading to diagnosis should be disregarded, as they cannot confer any preventive effect. Many case–control studies exclude a period of 6–12 months before diagnosis in order to rule out diagnostic tests. This can cause an unpredictably directed bias of its own (Weiss 1998, Zappa and Ciatto 2000).

Many methodological problems are reduced in settings wherein there is a centralised (organised) screening programme with active invitation at fixed dates; access to complete registers of cancers, invitations, and screening details; and a well-established (i.e., not recently initiated) screening programme (Zappa and Ciatto 2000). In this situation, tests done because of symptoms are not mistaken for screening tests, there is no recollection bias, and screening will have been available during the pre-invasive phase, when screening for cervical lesions can have a preventive impact.

Even in optimal conditions, there remains, however, the issue of self-selection bias. This arises from the response to invitation being an active choice influenced by personal characteristics, some of which may also influence the risk of cancer. Subjects are in this case self-selected into participants and non-participants with different baseline risks, regardless of any effect of the exposure itself. Self-selection bias is notoriously difficult to quantify in established long-running screening programmes with no easily available reference population without exposure to screening invitation. Collection of potential confounder information related to lifestyle and socio-economic status has been used to control this bias, but the resulting adjustments have usually been small, and in any case these data are typically not available in register-based studies (Kasinpila et al. 2011). Another approach has been to measure the risk ratio for other types of cancers and assume it to indicate the magnitude of the self-selection bias also affecting the risk of cervical cancer (Aklimunnessa et al. 2006). Unfortunately, there is no evidence of the validity of this assumption. There are some indications of the magnitude of this bias from early cohort studies wherein unscreened populations were used for comparison. When organised screening programmes were started, it was possible to draw expected cancer rates from the time before screening; doing this, Fidler et al. (1968) found a relative risk of 1.08 for those unscreened in the programme even though it is un-

clear how active their self-selection was – i.e., whether there was equal opportunity to participate. On the basis of early data from the Finnish screening programme, a relative risk of 1.6 was observed for those unscreened in the invitational programme (Hakama and Räsänen-Virtanen 1976). A third cohort study utilised expected cancer incidence rates drawn from neighbouring regions without screening programmes for a relative risk of 1.61 for invited non-participants (Magnus et al. 1987).

6.4.2 CASE–CONTROL STUDIES OF EFFECTIVENESS

Previous summaries of case–control evaluations of cervical cancer screening include the IARC collaborative study (IARC 1986) that consisted of a collection of studies assessing the association of cervical cancer risk and negative results in cervical cytology tests from Europe and North America. At least two later summaries are available (Zappa and Ciatto 2000, IARC 2005). This section of the chapter is based on a search of the literature for case–control evaluations of cervical screening published in 1990 or later with exposure based on both negative and positive smears. Sixteen such studies were identified and are listed in Table 4.

The first case–control evaluation published in this period included women under 75 with a diagnosis of cervical cancer in 1982–1985 and age-matched controls from Florence, Italy (Palli et al. 1990). Exposure was determined via interview of subjects and, in cases of death, through relatives whenever possible. Exposure was defined as ever versus never, excluding the last six months before diagnosis. Reported crude OR was 0.20. After adjustment for socio-economic factors, the OR estimate was 0.15. Recollection bias could be an issue with this study.

A study from Scotland determined the screening exposure prior to diagnosis and prior to smears leading to diagnosis of cervical cancer cases and deaths, and screening-register-based controls (Macgregor et al. 1994). Controls were matched for age and for having a negative smear at the date of presentation of the case. An OR of 0.67 was seen for incidence of screening exposure within three years. Effect on mortality was higher but was not defined for a specified window of exposure.

Table 4: Case–control studies of screening effectiveness published since 1990

First author	Year	Country/region	Case number, source, exclusions	Dates of diagnosis	Control number and selection	Screening history source and definition	OR for age at diagnosis if applicable (95% CI)
Palli	1990	Italy/Florence	191 ICCs, cancer registry, age under 75	1982–1985	540, resident list, matched for age and district of residence	Screening register, any smear up to 6 months before diagnosis	0.15 (0.09–0.25)
Herrero	1992	Bogota, Colombia; Costa Rica; Mexico City, Mexico; and Panama	759 ICCs, several hospitals, age under 70	1986–1987	1,430 age-matched controls from hospitals and census listings	Interview, last smear (excluding those within 12 months of diagnosis or interview)	0.40 (0.31–0.48)
Macgregor	1994	Scotland	282 SCCs from screening register	1982–1991	564, screening register, matched by age and by negative smear at time of diagnosis	Screening records, screened women had a negative smear before presentation	Negative smear within 3 years prior to presentation, 0.67 (0.39–1.14)
Macgregor	1994	Scotland	108 SCCs from screening register	1982–1991	216, screening register, matched by age and by negative smear at time of diagnosis	Screening records, screened women had a negative smear before presentation	Negative smear within 5 years prior to presentation , [0.25]
Sasieni	1996	UK / self-selected centres	294 ICCs IB+, pathology records, age below 70	1992	571 age-matched controls from local health authority registers	Screening register, result of last smear within 5 years and up to 6 months before	[0.52]
Hernandez-Avila	1998	Mexico / Mexico City	397 ICCs, eight hospitals, age below 75	1990–1992	1,005 controls of age 20–80 from National Household Sampling Frame	Interview, ever tested, last 12 months omitted, smears due to symptoms separated	0.38 (0.28–0.52)
Nieminen	1999	Finland / Helsinki University Hospital catchment area	147 ICCs, hospital, alive in 1994	1987–1994	1,098 controls, population register	Questionnaire, ever participated in organised or spontaneous screening	Organised: 0.36 (0.25–0.53) spontaneous: 0.73 (0.49–1.07)
Andersson-Ellström	2000	Sweden/Värmland	112 ICCs, pathology records	1990–1997	112 age-matched controls from population register	Screening register, last year omitted, screened within 3 or 6 years	Extracted ORs (within 3 y) all ages: [0.81] 20–59: [0.72]
Hoffman	2003	South Africa / Western Cape	524 ICCs IB+, two tertiary hospitals, age below 60	?	1,540 hospital based controls, series matched for decade of age, race, residence and hospital	Interview, time since last Pap smear, previous year omitted	Screened within 5 years compared to not screened, 0.3 (0.2–0.4)
Sasieni	2003	UK / selected centres	1,305 ICCs IB+, pathology records, age 20–69	1990–2001	2,532 age-matched controls registered with GPs in the same region	Screening register and laboratory records, time since last adequate smear, first 6 months excluded	Screened 6–42 months before diagnosis 20–39: 0.35–0.77* 40–54: 0.22–0.38 55–69: 0.18–0.42

* Range limits are extreme point estimates for a screening test within a one-year period in the 6–42 months before diagnosis.

Table 4 (Cont.)

First author	Year	Country/region	Case number, source, exclusions	Dates of diagnosis	Control number and selection	Screening history source and definition	OR for age at diagnosis if applicable (95% CI)
Zappa	2004	Italy/Tuscany	208 ICCs IB+, cancer register, age below 70 and resident for at least 5 years	1994–1999	832 age-matched controls from population register, resident for 5 years	Screening database, 12 month exclusion before diagnosis, any smear within 3 years / within 3–6 years / more than 6 years	Screened 1–3 years before diagnosis under 40: 0.35 (0.13–0.95) 40 + : 0.22 (0.12–0.42)
Crocetti	2007	Italy/Trento	61 ICCs IB+, cancer register, age 25–74, resident for at least 5 years	1995–2000	244 age-matched controls from health-service archive, resident for 5 years	Screening register, 12 month exclusion before diagnosis, any smear within 3 years/more than 3 years	Screened 1–3 years before diagnosis under 40: 0.50 (0.70–3.62) 40–59: 0.04 (0.01–0.32) 60+: 0.11 (0.01–0.95)
Andrae	2008	Sweden	1,230 ICCs, cancer register	1999–2001	6,124 age-matched controls from population register	National screening register, any smear 6-78 months prior to diagnosis	21–29: 0.42 (0.24–0.74) 30–65: 0.40 (0.34–0.47) 66+: 0.36 (0.24–0.53)
Ejersbo	2008	Denmark/Fyn	67 cases of ICC from cause-of-death register	1997–2001	67 living at date of diagnosis from population register, matched for age and place of residence	Screening register, smears within 5 years, excluding diagnostic smears a few months before diagnosis	Under 60: [0.25] 60+ : [0.87] overall: [0.54]
Yang	2008	Australia / New South Wales	877 ICCs, age 20–69, cancer registry	2000–2003	2,614, from screening register, matched for age	Screening register, any smear within 3–48 months of diagnosis	20–29: 0.10–0.24† 30–49: 0.05–0.19 50–69: 0.03–0.10
Sasieni	2009	UK / selected centres	4,012 ICCs, age 20–69, histology laboratory records	1990–2008	7,889 population-based controls matched for age and region of residence	Screening records, smear in a 3-year period before dg in the next 5-year window	25–29: 1.11 (0.89–1.50) 35–39: 0.55 (0.44–0.69) 45–49: 0.37 (0.29–0.48) 55–59: 0.26 (0.19–0.36)
Kasinpila	2011	Thailand / Khon Kaen	130 ICCs, age 30–64, treated at 4 hospitals	2009	130 hospital and 130 companion controls matched for age	Interview, most recent test within 6, 12, 36 or 36+ months	0.27 (0.13–0.57), screened 12–35 months before diagnosis

† Range limits are point estimates of one or two smears within the window of exposure.

All 179 cases of invasive cervical cancer treated at a Finnish hospital in 1987–1994 and alive in 1994 were classified according to their exposure to organised screening, spontaneous screening, and gynaecological visits (Nieminen et al. 1999). Controls from the hospital catchment area were sampled from the population register, along with socio-economic data for both cases and controls. Exposure and risk factors were derived through questionnaire. The exposure window was closed the year before diagnosis. The ORs for ever- versus never-screened were adjusted for 10-year age groups. The ORs were also adjusted for the other two types of exposure. An OR of 0.38 was reported for organised screening, while the effect estimate was smaller for spontaneous screening (OR: 0.82). There is a possibility of exposure misclassification and selection bias, although this should not affect the striking difference in effect seen between the two modalities of screening. Cases were restricted to those patients who had not died from their cancer, and it is possible that effect estimates for this population are smaller than among all diagnosed cases.

A Swedish study included all 112 cervical carcinoma cases diagnosed at a hospital in Karlstad (Andersson-Ellström et al. 2000). The screening histories were derived from cytology registers and compared with population-based controls matched by age. A Chi-square test was used to compare proportions of smears 1–3 or 3–6 years prior to diagnosis. The extracted crude OR was 0.81 over all ages for screening within three years.

In countries with a less developed health infrastructure, evaluation results can nevertheless be promising and long-lasting risk reductions observed, but these results should be interpreted with caution. In South Africa, a hospital-based study with 524 stage-IB+ ICCs and series-matched controls found that the association of Pap smears more than 15 years previously between the risk of cervical cancer was significant (OR: 0.5, 95% CI: 0.4–0.7) (Hoffman et al. 2003). In this study, data on several risk factors were collected and adjusted for by means of unconditional multiple logistic regression in order to limit the effects of self-selection bias, but the magnitude and duration of the effect of screening suggest that this was probably not entirely successful. In addition, screening history was based on interview, wherein recollections of events decades past were used as exposure information. A real risk of misclassification bias is present. The authors refer to a publication in Latin America that compared screening histories for newly diagnosed cases of cervical cancer in four hospitals in four countries with those of age-matched hospital or community controls (Herrero et al. 1992), as an example of good effect in developing countries. The Latin American study excluded 12 months prior to diagnosis and found the protective effect to increase with age from an OR of 0.56 for ages under 30 at diagnosis to 0.25 for ages 40–49. Similar risk reductions were observed for adenocarcinoma (OR: 0.50), which could indicate a significant com-

ponent of bias in the estimates. Exposure was defined by interview, which could lead to misclassification, and selection bias may be a major issue. Another study, with 397 invasive cases diagnosed at six hospitals in Mexico City and population-based controls stratified by age, also used interviews to determine ever-versus-never screening status with exclusion of the last year before diagnosis (Hernandéz-Avila et al. 1998). Estimates were adjusted for age, socio-economic status, and sexual history. When diagnostic smears were factored out, an OR of 0.38 was observed (without this distinction, the OR was 0.76).

Sasieni et al. (2003) investigated primarily the length of the protective – or, in fact, selective – effect of a negative screening test with different age groups but reported also conditional logistic regression results for having an adequate screening test within specific one-year intervals before diagnosis in different age bands, ignoring all smears within six months of diagnosis. Screen-detected cancers could not be identified in this study, but all microinvasive carcinomas were excluded. Cases were diagnosed in 1990–2001. Compared with those without any adequate screens registered, the ORs of cervical cancer for those 20–39-year-olds with their last but not only screen 0.5–1.5, 1.5–2.5, and 2.5–3.5 years before diagnosis varied between 0.35 (0.23–0.54) and 0.77 (0.53–1.11). For 40- to 54-year-olds, the ORs varied between 0.22 (0.14–0.34) and 0.38 (0.26–0.54) and for 55–69-year-old women between 0.18 (0.11–0.30) and 0.42 (0.27–0.65). An earlier UK study defined screening histories for cases of cancer diagnosed in 1992 in participating districts and for age-matched community-based controls (Sasieni et al. 1996). The main result again was risk reduction by time after negative smear, but all smears were reported such that a crude OR of 0.52 can be extracted for stage-1B+ cancers. Screening registers were used to determine exposure six months to five years before diagnosis for women under 70.

In Trento, Italy, age- and morphology-specific case–control analysis produced high effect estimates for adenocarcinoma also (Crocetti et al. 2007). The study identified all cervical cancers diagnosed at the ages of 25–74 in 1995–2000 but excluded microinvasive carcinomas and cases among women resident in the region for less than five years. Controls were matched by age to their cases, had to be alive at the time of diagnosis, and also were residents for the past five years. Screening tests done less than 12 months before diagnosis were excluded from exposure. High effect estimates were found for tests performed 1–3 years before diagnosis. For example, an effect of 96% was observed for invasive cancer in the age range 40–59. The overall protective effect for adenocarcinoma too was high (OR: 0.24). The effect of screening women under 40 was smaller and non-significant (OR: 0.50, 95% CI: 0.70–3.62). Possible biases affecting the findings of this study are caused by the exclusion of 12 months' screening history in a relatively short exposure window (three years) and self-selection.

The same selection criteria for cases, controls, and screening history were used in another Italian study (Zappa et al. 2004). Here the effect on the risk of SCC was of the same magnitude (OR: 0.15) for women 40 or over, but a clear effect was also observed for women under 40 (OR: 0.16). Effect estimates for adenocarcinoma were not significant. Estimates were adjusted for civil status and place of birth, but crude estimates were not reported. The authors discuss the possibility of residual bias but again argue that comparisons of age groups and histological types of cancers should not be affected even if absolute effects are overestimated.

A register-based study from Odense, Denmark, analysed the screening histories prior to diagnosis of 67 women who died of cervical cancer and 67 population-based age-matched controls (Ejersbo 2008). Dichotomous screening history determinants as to ever or never screened, more than two screening tests recorded, and any screening tests within five years of diagnosis were reported. Diagnostic smears were excluded, with this classification based on hospital records. Results were presented as numbers and proportions, so the odds ratios in Table 4 are crude odds extracted from the available counts. Estimates were different in the two age groups: women up to 59 and 60 or over. The odds ratios were 0.25 for the younger women and 0.87 for the older women, with an overall OR of 0.54.

An Australian study from New South Wales evaluated the association with screening tests four years before diagnosis (Yang et al. 2008). Tests performed within three months of diagnosis were excluded from the screening history. Exposure was categorised into no screening, irregular screening (that is, one screening event within the window of exposure), and regular screening (biennial, as recommended). Very high protective effects for all invasive cervical cancers were observed, and odds ratios did not vary by age. However, controls were drawn from the screening register, which may introduce a strong bias in favour of screening if coverage is not complete. Furthermore, protective-effect estimates for non-squamous cancers were also very high (up to 93%) and did not differ significantly from those for squamous carcinomas. This may be an indication of a large selection-bias component.

A very large study from the UK, with 4,012 women aged 20–69, investigated the effect of screening participation in overlapping three-year bands on the cervical cancer risk in the following five-year period (Sasieni et al. 2009b). Protective effects of 60–80% were estimated for cervical cancer in women at ages of 40 or more, but there was considerable age-dependency of effectiveness, with screening under the age of 25 having little or no effect. The main potential bias remaining in this study is self-selection bias. For the difference in effect for different ages to be invalid, however, the selection bias would have to be age-specific. The authors argue that this is unlikely and unsupported by evidence.

A nationwide cervical cancer screening programme with recommended every-five-year cytology screening tests for women of ages 35–60 was started in 2005

in Thailand. The effectiveness of the programme was evaluated by investigation of screening histories of 130 (out of 135 invited to participate in the study) women with ICC, treated in four clinics in north-east Thailand (Kasinpila et al. 2011). Two control groups were recruited, with frequency-matching by age: 130 hospital controls from other wards and 130 presumably healthy hospital companion controls. There were no significant differences in risk factors or screening history between the two control groups, so they were combined for the final analysis. Data on risk factors known to be associated with both screening history and cancer were collected by interview. Socio-economic markers were strongly associated with cancer risk, but in a multivariate analysis the only significant variables were alcohol consumption, age at first intercourse, and the use of oral contraception. Adjustment for these variables did not make a large difference for the effect estimates. The effect of screening 6–11 months before diagnosis was estimated at 1.38, which may indicate that there were still diagnostic tests within this period. Screening 1–2 full years before diagnosis was associated with an adjusted relative risk of 0.27, and the relative risk was 0.42 with an interval of three years or more. However, as the study was based on interview, there is a risk of recall and misclassification bias. Tests done because of symptoms and those leading to the detection of cancer could not be directly identified in this study, but the final analysis of screening history did exclude those smears taken within six months of diagnosis. This approach is widely used but carries a risk of overestimating the screening effect (Weiss 1998). The authors discuss the possibility of residual confounding and, especially, the potential effect of unmeasured self-selection bias and conclude that the effect estimates probably are too optimistic.

Earlier studies often defined screening exposure in terms of ever- versus never-screened. All studies done after 2000 define screening history in a specific window of exposure, usually related to the recommended screening interval (but not necessarily coinciding with the duration of protection). In general, later studies have also included more detailed descriptions of the methodology employed. A gradual shift over time toward conditional logistic regression was noted in the statistical analysis methods. Many studies with large effects were vulnerable to misclassification of exposure due to a lack of administrative registers, and also the inclusion of all relevant outcome events may have been suboptimal. Most studies did not demonstrate good means of discriminating between diagnostic and screening tests, or screen-detected and symptomatic cancers, and had to resort to restriction of the window of exposure, which brings a risk of bias. None of the studies had the ability to correct for selfselection bias directly; adjustment for risk factors was employed in some instances, but the effect on estimates was small when reported.

6.5 ADVANCES IN BIOTECHNOLOGY – IMPROVED TEST SENSITIVITY AND PRIMARY PREVENTION

6.5.1 HPV TESTING

Despite the impressive track record of cytology, some features have been objects of criticism. The sensitivity of a single smear for intraepithelial lesions can be low and variable. Cross-sectional sensitivities ranging from 30% to 87% have been reported for a test threshold of LSIL and a disease threshold of CIN1, and from 44% to 99% for a disease threshold of CIN2 with specificities of 91% to 98% for CIN2+ (Nanda et al. 2000). However, in practice, repeated smears during the long period of preclinical lesion development ensure that compounded programme sensitivity is higher. The compounded programme sensitivity benefits not only from several tests being done over the lifetime of preclinical lesions but also from the fact that some of the lesions missed during one screening episode will regress in the interval before the next screening round. Another important and related drawback is the poor inter-observer and also intra-observer reproducibility of results, which is due to the subjective evaluation of cytopathology, the quality of which, in turn, is highly dependent on the level of training and experience of the laboratory staff (Stoler and Schiffman 2001).

Screening based on primary HPV DNA testing can offer improved performance in these areas: HPV testing has a relatively high sensitivity for precancerous lesions; good reproducibility of test results; and, naturally, also a high specificity for HPV disease (Ronco et al. 2010a). On the other hand, specificity for higher-grade lesions (CIN2+, CIN3+) is lower. This becomes a problem at younger ages, when HPV infections are prevalent but progression to cervical cancer is rare. For this reason, primary HPV DNA testing can only be recommended for women aged 30 and above (Schiffman et al. 2010).

HPV testing can be based on the detection of viral DNA, mRNA, or viral proteins. Most screening programmes use DNA-based diagnostic test systems. There are two main types of tests for the detection of HPV DNA. The most widely used tests are nucleic acid hybridisation assays for collective detection of 13 high-risk types of HPV viruses (Cuzick et al. 2008). In addition, tests based on the polymerase chain reaction are available (Gravitt et al. 2008). RNA-based tests are an attempt to increase the specificity of HPV tests by detecting the presence of transcripted messenger RNA of viral oncoproteins. The rationale is that active oncoprotein production in infected cells is more prognostic of progressive neoplasia than is the presence of viral DNA. The main oncoprotein mRNAs of interest are E6 and E7 (Andersson et al. 2012).

Another approach is the detection of tumour suppressor gene products (Brown et al. 2012). Normally the cell cycle is carefully regulated, and mitosis usually in-

creases the expression of the p16 gene, which, in turn, represses further cell replication. The E7 HPV oncoprotein inhibits the function of the p16 gene product, which leads to over-expression of p16. Immunocytochemical detection of the resulting high concentrations of the p16 product is possible either in cytological or in histological samples, but the method still suffers from non-standardised interpretation and poor reproducibility (Tsoumpou et al. 2009, von Knebel Doeberitz et al. 2012).

There are also methods available for direct detection of the viral oncoproteins E6 and E7, but there is little experience or information available with respect to the performance of these tests in screening applications (Laurenson et al. 2011).

HPV screening based on the detection of DNA is a viable alternative in low-resource settings if the laboratory infrastructure and experience needed for reliable cytology are lacking (Sankaranarayanan et al. 2009). With automated analysis of HPV DNA tests, it is easier to achieve high and consistent performance. HPV tests may bring improvements even in settings with good-quality cytology, thanks to increased sensitivity and a long-lasting NPV, but this requires that the screening protocol be appropriately chosen and that the benefit-to-harm ratio be rigorously evaluated and monitored (de Kok et al. 2012).

6.5.2 VACCINES

There are currently two commercial prophylactic HPV vaccines available, and both are used in national vaccination programmes. They confer high immunity against new infection with HPV types 16 and 18, and one contains, in addition, antigens from low-risk types 6 and 11, which are the most important causative agents for genital condyloma (Einstein et al. 2011).

Prophylactic vaccines may have the potential to answer the challenge of possible lower screening effectiveness in the younger age groups, under 30 (Brotherton et al. 2011). When the vaccinated cohorts reach the core screening ages, 30–64, there will probably also be impacts on the optimal configuration of the screening programmes. Vigilant monitoring and evaluation is, therefore, especially vital in the coming years, because of these near-future challenges and opportunities of cervical cancer prevention. Vaccinated and unvaccinated cohorts should be audited separately, since there may be a need to provide separate screening programme protocols according to vaccination status (Lehtinen et al. 2012).

Therapeutic vaccines are under development, and these may affect treatment options and follow-up (Barrios and Celis 2012). If these vaccines have long-lasting immunological effects, there may also be consequences for the optimal screening algorithm for treated women.

7 AIMS OF THE STUDY

The aim of this study was to evaluate performance and age-specific effectiveness of cervical cancer screening, by focusing on audit studies examining the Finnish cervical cancer screening programme within case–control designs wherein information on the outcome of screening was taken into account.

The study also developed further quality assurance protocols for integration into the programme. More detail-level research problems included assessment of the quality of the screening monitoring database, auditing of test performance and validity in the screening programme, and evaluation and auditing of the effects of the cervical cancer screening programme on cervical cancer in an agespecific manner. Specific studies included the following:

1. Evaluation of the completeness and accuracy of outcomes (cervical precancerous lesions and cancer) recorded in the cervical cancer screening register, by comparing data with both the cancer register and the administrative hospital discharge register

2. Evaluation of the performance and validity – specifically, the false-negative rate, sensitivity, specificity, and reproducibility – of the cytological screening test in an outcome-based audit of cytology within the programme

3. Evaluation of variation in cross-sectional performance indicators by screening laboratory and the assessment of any association with longitudinal sensitivity

4. Audit of screening histories of cervical cancers and controls so as to identify areas of potential for improvement within the screening policy and service. Particular emphasis was given to evaluating the effect of one screening episode at a given age in the programme against cervical cancer and cervical cancer death. Any variation in effectiveness across age at invitation was studied, with particular attention to the marginal age groups of 25 and 65, which are currently not recommended but merit further consideration in terms of whether they should be targeted by the programme

The output from these evaluations is essential for producing information needed to support decision-making on programme modifications and for developing feedback to the screening service providers.

8 MATERIALS AND METHODS

8.1 THE CERVICAL CANCER SCREENING PROGRAMME IN FINLAND

Cervical cancer screening was introduced in 1963 as a regional pilot programme and expanded to national coverage by 1970 (Anttila and Nieminen 2000, Anttila and Nieminen 2007). The Government Decree on Screenings (1339/2006) specifies that municipalities shall organise fiveyearly screening tests for cervical cancer, starting at the age of 30, with the last invitation issued at 60. All women in the targeted birth cohorts are sent a personal invitation to be screened. Nearly complete coverage of the recommended target age groups by invitations has been reached in recent years (see the unpublished results in Table 5). The coverage showed some regional variation in the '90s, and a number of municipalities have also invited 25- or 65-year-old women.

Table 5: Invitation coverage of the cervical screening programme in Finland in 1990–2009 (%)

	Invitation year							
Age	1990–1994	1995–1999	2000–2004	2005	2006	2007	2008	2009
25	40	32	36	35	28	30	29	15
30	70	74	89	93	92	94	98	98
35	92	92	98	98	98	99	99	100
40	98	99	99	99	99	100	100	100
45	98	95	97	99	99	100	100	100
50	98	99	100	100	99	100	100	100
55	79	91	97	98	96	98	100	100
60	59	76	88	95	94	97	99	99
65	12	15	16	16	15	16	16	3

Data source: Cervical cancer screening register at the Mass Screening Registry.

The screening test in use is conventional cytology, except in a number of municipalities in Southern Finland that participate in a randomised public-health trial comparing primary HPV DNA screening to conventional cytology (Anttila et al. 2006, Leinonen et al. 2009, Anttila et al. 2010). Borderline cytology (ASC-US, AGC-NOS according to Bethesda terminology (Solomon et al. 2002), or equivalent Pap II findings before year 2006) triggers a new, so-called risk-group invitation in

12 months instead of the five years in the default programme for screen negatives. LSIL or worse (LSIL+), or Pap III–V (before 2006), leads to immediate referral for colposcopy (Nieminen et al. 2010). Women with referral are actively followed, and data on the histological confirmation are collected by the screening laboratory and eventually registered in the screening database of the Mass Screening Registry.

8.2 DATA SOURCES

Monitoring and evaluation of screening, as of other public-health interventions, are actions greatly facilitated by high-quality registers and the ability to link them at the level of the individual, preferably by means of a personal identifier. The Nordic countries, Finland among them, have a long history of population registration and national registers for several aspects of health care, such as cancer screenings, birth, prescriptions, hospital discharge, health insurance payment, cardiovascular disease, and infectious disease. These can be used to explore associations of risk factors or interventions and outcomes at a low cost of additional resources. This study has utilised the cervical cancer screening database, the cancer register, and the administrative hospital discharge register as data sources.

8.2.1 THE SCREENING REGISTER

Complete, or nearly complete, individual-level data on the cervical cancer screening programme are currently available in the electronic database from 1963 to 1976 and again from 1990 onward. For this study, data from 1990 to 2009 were used to construct screening histories for the study subjects. Relevant data included invitation year, inviting municipality, type of invitation (age- or risk-based), screening laboratory, date and result of screening test, any follow-up recommendation, and date and result of histological verification.

8.2.2 THE CANCER REGISTER

The histological verification, or diagnosis, data registered in the screening register are based mainly on the first colposcopy and biopsy after referral. For the cancer register, on the other hand, all cancer notifications from hospitals, physicians, and pathological and haematological laboratories and death certificates with cancer as a cause of death are reviewed before a summary, or synthesis, of all available information is registered. The result is a very complete and accurate register of

malignancies (Teppo et al. 1994). In addition to malignant neoplasms, the cancer register includes some premalignant conditions of the cervix uteri (the cervix uteri is represented by topography codes C53.0–C53.9 in the ICD-O-3 classification (Percy et al. 2000)). These are dysplasia gravis, which is not specified in ICD-O-3 and therefore is registered with an in-house code; CIN3 with ICD-O-3 morphology code 8077/2, SCC *in situ* (CIS) (8070/2); carcinoma *in situ* NOS (8010/2); and AIS (8140/2). The classification of lesions into these categories has developed over time, and also there was clear under-registration of premalignant diagnoses until the late 1990s (Finnish Cancer Registry 2009).

Invasive cervical cancer can be separated into more than 40 morphological types, according to the WHO (IARC 2005). These can be grouped into SCCs, adenocarcinomas, and other specified and unspecified malignancies. The overwhelming majority of cervical malignancies derive from the epithelium, or, in other words, are carcinomas. In addition, a small number of sarcomas are located at the cervix. The SCCs are the most common type of cervical cancer, and they are also the type most effectively prevented by cytology screening (*ibid*.). Of all 1,548 cervical cancers diagnosed in Finland in 2000–2009, 62% were SCCs, 29% adenocarcinomas, 4% other specified or unspecified carcinomas, 2% sarcomas, and 3% unspecified malignancies (see Table 6, from Paper IV and unpublished work). Out of the 545 deaths from cervical cancers in 2000–2009, 61% were due to SCCs, 28% adenocarcinomas, 5% other or unspecified carcinomas, and 1% sarcomas, and 5% were caused by morphologically unspecified cervical malignancies.

The stage of disease registered in the FCR represents summary information derived from all cancer notifications, whether clinical or pathological, and is divided into localised, locally spread, metastasised, and spread but to an unknown extent. This staging is applied to all cancer sites. However, the recommended staging for cervical cancer according to FIGO (Pecorelli et al. 2009) is the clinical stage at presentation, which represents the pre-surgical information that is available when treatment is planned. The clinical stage does not have the same predictive value with respect to survival as post-surgical, or pathological, staging, but it is nevertheless favoured because of comparability issues. We determined the clinical stage by going through the cancer notifications for cervical cancers diagnosed (IV) or cause of death (V) in 2000–2009 and were able to derive a clinical stage for 66% of these cases. Absence of clinical stage was complemented by pathological stage information in 23% of cases, and a further 11% were left with an unknown stage.

Table 6: Morphology of malignant tumours of the uterine cervix diagnosed or causing death in Finland in 2000–2009 (source of case material in papers IV and V)

	incident cancers	%	cause of death	%
Total	**1548**	***100.0***	**545**	***100.0***
Squamous-cell carcinomas	**967**	***62.5***	**333**	***61.1***
8052 Papillary squamous-cell carcinoma	1	*0.1*	0	*0.0*
8070 Squamous-cell carcinoma, NOS	746	*48.2*	328	*60.2*
8072 Squamous large-cell, nonkeratinising carcinoma	1	*0.1*	0	*0.0*
8076 Squamous-cell microinvasive carcinoma	217	*14.0*	4	*0.7*
8083 Basaloid squamous-cell carcinoma	1	*0.1*	1	*0.2*
8084 Clear-cell squamous carcinoma	1	*0.1*	0	*0.0*
Adenocarcinomas	**453**	***29.3***	**151**	***27.7***
8140 Adenocarcinoma, NOS	402	*26.0*	139	*25.5*
8260 Papillary adenocarcinoma, NOS	3	*0.2*	1	*0.2*
8310 Clear-cell adenocarcinoma, NOS	2	*0.1*	0	*0.0*
8380 Endometrioid adenocarcinoma, NOS	35	*2.3*	7	*1.3*
8384 Adenocarcinoma, endocervical type	4	*0.3*	0	*0.0*
8441 Serous cystadenocarcinoma, NOS	2	*0.1*	1	*0.2*
8460 Papillary serous cystadenocarcinoma	1	*0.1*	1	*0.2*
8480 Mucinous adenocarcinoma	2	*0.1*	2	*0.4*
9110 Mesonephroma	2	*0.1*	0	*0.0*
Other specified epithelial tumours	**37**	***2.4***	**11**	***2.0***
8013 Large-cell neuroendocrine carcinoma	1	*0.1*	0	*0.0*
8015 Glassy-cell carcinoma	2	*0.1*	1	*0.2*
8041 Small-cell carcinoma	6	*0.4*	4	*0.7*
8098 Adenoid basal carcinoma	3	*0.2*	0	*0.0*
8200 Adenoid cystic carcinoma	3	*0.2*	1	*0.2*
8246 Neuroendocrine carcinoma, NOS	6	*0.4*	0	*0.0*
8560 Adenosquamous carcinoma	16	*1.0*	5	*0.9*
Epithelial tumour, NOS	**21**	***1.4***	**16**	***2.9***
8010 Unclassified carcinoma	21	*1.4*	16	*2.9*
Other specified tumours	**23**	***1.5***	**4**	***0.7***
8800 Sarcoma, NOS	2	*0.1*	1	*0.2*
8890 Leiomyosarcoma, NOS	5	*0.3*	0	*0.0*
8910 Embryonal rhabdomyosarcoma, NOS	2	*0.1*	0	*0.0*
8933 Adenosarcoma	1	*0.1*	0	*0.0*
8935 Stromal sarcoma, NOS	3	*0.2*	1	*0.2*
8950 Mullerian mixed tumour	1	*0.1*	0	*0.0*
8980 Carcinosarcoma, NOS	6	*0.4*	1	*0.2*
8990 Mesenchymoma	2	*0.1*	1	*0.2*
9100 Choriosarcoma, NOS	1	*0.1*	0	*0.0*
Other neoplasm, NOS	**47**	***3.0***	**28**	***5.1***
8000 Unclassified malignancy	47	*3.0*	28	*5.1*

8.2.3 THE HOSPITAL DISCHARGE REGISTER

Since 1996, ICD-10 diagnosis codes have been used to register inpatient episodes and day-surgical procedures in the HDR. From 1998 onward, all outpatient visits in the public sector have been registered. Data include visit diagnosis, treatments given, and date of visit or discharge. The diagnoses found in this register usually cannot be used as incidence data without further information generated, for example through linkage to other health-care registers.

8.3 LINKAGE OF DATA SOURCES

8.3.1 LINKING REGISTERS OF CERVICAL LESIONS (I)

Accurate linkage of register data is possible and relatively simple because of the personal identifier given to all residents of Finland. For comparison of the histological diagnoses found in the various data sources, we chose a time period when also the HDR and the cancer register could be expected to have good coverage of precancerous lesions – i.e., 1998 onwards. We started with the population of women screened in 1998–2007 who had received a referral for colposcopy because of a positive screening test and therefore should have a histologically confirmed diagnosis in the screening register. In the years examined, there were 16,353 referrals, involving 15,912 women. The confirmed histological diagnoses resulting from these referrals included 4,309 cases of CIN1–2, 2,152 cases of CIN3/AIS, and 185 cases of ICC. The referrals were linked to the HDR, which yielded 54,263 cervical, vulvar, vaginal, and uteral health-care contact episodes, involving 12,832 women. For 2,644 of these women, an invasive or pre-invasive diagnosis was found in the FCR. We were interested especially in the completeness and accuracy of screening registration and had to define the diagnoses found in the other two registers by the mode of detection in relation to screening. We used the criterion of diagnosis within 12 months of a positive screening test (i.e., a screening test with a referral for histological confirmation) for defining a diagnosis as screendetected, and hence eligible for inclusion in the screening records. Whenever multiple diagnoses were found within this time window in a register, the highest-grade cervical neoplastic lesion or cancer was chosen.

The diagnosis in the cancer register was regarded as the most accurate available, because of multiple notifications and the verification procedures in use. However, in the 10 cases wherein an invasive cancer was suggested by either of the other two registers and no record for the woman existed in the cancer register, a thorough

review of the patient history records was done. The review resulted in updating of the cancer register with two additional ICC diagnoses and a change of primary site from an undetermined female genital cancer to specifically the cervix in one instance. The updated cancer register records on ICC were used as reference in the subsequent comparisons of registers.

8.3.2 POTENTIAL FALSE-NEGATIVE SCREENING TESTS (II AND III)

For auditing of screening-test performance, seven screening laboratories active during the 1990s were recruited and the data in the screening register corresponding to the active time for each laboratory were linked with the cancer register files for CIN3/AIS and ICC in 1990–1999. The screening records included in the linkage amounted to 953,610 screen samples. Audit cases were defined as CIN3+ diagnoses preceded by a Pap I or Pap II screening test that did not result in a referral. These tests were regarded as potentially false negative and were included in the review phase.

8.3.3 CANCER AUDIT OF SCREENING HISTORIES WITH CONTROLS (IV AND V)

For the evaluation of effectiveness, all ICC cases diagnosed in 2000–2009 and all deaths due to cervical cancer in 2000–2009 were age-matched by birth month and year to six controls from the population register. Women had to be alive and not diagnosed with cervical cancer at the time of diagnosis of their matched case. Two incident-case women had data restriction in place and were excluded from the screening history and case–control analysis for a final incidence-case count of 1,546. The cases and the controls were then linked to the screening register data from 1990–2009. There were 39 cases of death with index invitation by age before 1990. These cases and their controls were excluded from the screening history analysis because we did not have data on the screening exposure. The final sum of cases of cervical cancer death in the screening history and case–control analysis was 506.

8.4 MODE OF DETECTION AND SCREENING HISTORY

8.4.1 MODE OF DETECTION (IV AND V)

The mode of detection with respect to screening is a variable that has relevance only for cases of cervical cancer, not controls. This type of classification has been recommended by the ENCR (ENCR 2001) for routine use by cancer registers for cancers targeted by screening, and an updated classification was used for the first time in the Finnish Cancer Registry for cervix and breast cancers in 2011 (Finnish Cancer Registry 2011). Cases can occur in the non-invited population, among nonparticipants, or among those participating in screening. Participants' cases are further separated into screen-detected and interval cases. Screen-detection was defined as the diagnosis of cancer within 12 months of a positive screening test with referral for colposcopy and biopsies. This definition is recommended by the European guidelines (Arbyn et al. 2008a) and is especially suitable for application in the Finnish screening programme because of the near-perfect compliance with the referral colposcopy (Ronco et al. 2009).

8.4.2 PREVENTIVE-SCREENING STATUS (IV AND V)

The item for preventive-screening status represents screening exposure and can therefore be addressed for cases and controls alike. For studying the effectiveness of screening tests, we used the outcome of the age-group invitational screening event immediately preceding the date of diagnosis of the case, or the corresponding date for the matched controls, as the determinant of exposure. Because of the default five-year interval in the Finnish screening programme, this index screening event was defined as the last age-group programme test within the 66 months preceding diagnosis. This period is the sum of the five-year screening interval and a six-month allowance for test date variation within the invitation year. Screening events leading to the detection of cancer were disregarded, as they do not contribute to the preventive impact of screening. Instead, the previous invitational screening event was used to determine the index-screening status of screen-detected cases and their matched controls, thus maintaining equal opportunity of exposure for cases and controls. The outcome of the last invitational screening event was indexed as negative when there was no further recommendations made and the default screening interval was applied, borderline when a recommendation for re-screening within a shorter interval was

made, and referral when there was a recommendation for colposcopy directly as a result of the age-group invitational smear.

8.5 REVIEW OF ARCHIVED SMEARS (II AND III)

Sensitivity failure of screening can be expressed in terms of false-negative tests or episodes. To this end, we used the rereading of archived smears that were classified as negative in the sense that no referral for further confirmation was made but that nevertheless preceded a diagnosis of CIN3+. There were 474 such case smears in the cytology audit (II and III). These smears, and two sequential smears for controls, were requested from the archives of participating laboratories, and 395 (83%) were retrieved, along with 787 controls. Laboratories then reanalysed their own smears, cases and controls, while blinded to the original reading and the case history. All smears also underwent blinded analysis by a reference laboratory at the Department of Gynaecology and Obstetrics of the Helsinki University Central Hospital. Whenever these two readings differed significantly from each other (207 case smears and 159 control smears), the smear was additionally analysed by an expert panel who then provided the best attainable, or gold standard, cytology. At the end of this process, each smear had four cytology results: the original result (either Pap I or Pap II), the laboratory reanalysis result, the reference laboratory's result, and the gold standard cytology. Any smears upgraded to positive cytology by the gold standard were considered clearly positive false negatives, while comparisons of the different readings were used for estimation of test validity and reproducibility.

8.6 STATISTICAL METHODS

8.6.1 VALIDATION OF THE REGISTER DIAGNOSIS (I)

Completeness of records was described by two measures. These were coverage, referring to the presence of any diagnosis wherein a neoplastic diagnosis was present according to the reference standard, and sensitivity, referring to the presence of a diagnosis at least as severe as that suggested by the reference standard. The cancer register was used as a reference standard for invasive disease, and for CIN3+, CIN2+, and CIN1+, the reference standard was the population of cases above each threshold of severity according to at least one of the three registers under comparison.

The PPV and the kappa coefficient were used to describe accuracy of register diagnosis. The PPV was only for ICC diagnoses only and corresponds to the proportion confirmed by the FCR. Linearly weighted kappa coefficients were calculated for pair-wise inter-rater agreement of specific register diagnoses (Gwet 2002).

8.6.2 CYTOLOGY AUDIT AND LABORATORY PERFORMANCE (II AND III)

False negativity was calculated with two cut-off points for both the original reading and the reference cytology. The smears included in the audit were originally Pap I or Pap II. The Pap II class includes borderline changes that can be indicative of dysplasia but also a large proportion of reactive, or benign, findings. We did not have the means to distinguish these groups, so the main results were presented with both Pap I and Pap II included as originally negative smears. Accordingly, the main cut-off for the gold-standard cytology used as reference was LSIL+, which generally requires a referral for colposcopy. As the reporting of cytology moved from a modified Papanicolaou classification to Bethesda 2001 terminology in 2006, the review was recorded in terms of Bethesda.

For test sensitivity, we used the results of the review of case and control slides in the original laboratory in comparison to the gold-standard cytology. We calculated proportions of smears correctly diagnosed at three thresholds or higher. The lowest threshold was ASC-US+, which is the minimum requirement for abnormal smears, as it will at least trigger the intensified screening protocol. The intermediate threshold was LSIL+, which usually triggers a referral for colposcopy, and the highest threshold used was HSIL+, representing a high risk of dysplasia that requires treatment and more urgent referral. The inclusion of controls in the review enabled us also to calculate specificity, which was equivalently reported as the proportion of smears correctly reported below the above-mentioned thresholds. Confidence limits for sensitivity and specificity were estimated at 95% confidence levels assuming a binomial distribution.

Inter-observer reproducibility of cytology was analysed by cross-tabulation of the original laboratory rereading results with the reference-laboratory results. Unweighted and linearly weighted kappa statistics were estimated both for specific cytological results and for results grouped into three categories by resulting recommendation (normal screening interval, intensified screening, and referral for colposcopy).

Performance indicators for each participating laboratory were tested for heterogeneity in logistic regression models with age and year as explanatory variables. For heterogeneity testing, we restricted the material to the years when all participating laboratories were active, in order to have equal opportunity for cases to arise in

follow-up. The performance indicators tested included the intensified-screening recommendation rate; the referral rate; detection rates of CIN1+, CIN2+, and CIN3+; and PPVs for CIN1+, CIN2+, and CIN3+. The audit CIN3+ case rate compared to total screening tests and the proportion of audit CIN3+ cases to all CIN3+ cases were used as outcome indicators. The heterogeneity across laboratories for the former was tested in a logistic regression model, and the heterogeneity of the latter was tested in a log-binomial regression model of age and year.

8.6.3 SCREENING EFFECTIVENESS (IV AND V)

Screening effectiveness was estimated in a case–control design with screening attendance in the index screening event as exposure. Conditional logistic regression was used to estimate the odds ratios for the association of attendance and either cervical cancer (IV) or cervical cancer death (V), with a date of diagnosis in the screening interval up to and including any screen-detected diagnoses from the following screening event. In order to account for selection bias – that is, the difference in baseline risk between those choosing to attend and those choosing not to attend – a self-selection factor (*Sf*) was utilised. The self-selection factor was estimated as the risk ratio (approximated by the odds ratio) of those choosing not to attend screening after invitation and those not invited within the matched case–control material. This factor was based on all cervical cancers (including microinvasive carcinomas) for correction of the ORs for the effects on incidence and FIGO stageIB+ cancers for correction of the ORs for the estimated effects on mortality. Multiplying the crude OR by *Sf* would yield an OR approximating the relative risk of disease in the participants as compared to those not invited, which is not an estimate of screening effect. The risk of all women not invited is higher than the risk of those who would participate if invited, so the correction is too small. Information on participation is needed also; hence, the corrected ORs for screening effect were generated through application of Formula 5:

$$OR_{corr} = \frac{p \times OR_{crude} \times Sf}{1 - (1 - p) \times Sf}$$

from the methodology paper by Duffy et al. (2002). This formula is applicable to self-selection corrections in screening evaluations and has previously been used in case–control studies of the effectiveness of breast cancer screening (Otto et al. 2012). The participation rate (p, 71%) of those invited in the screening programme in 1990–2009 was used in the correction.

The above formula estimates the effect as the risk in those participating compared to those who would participate if invited. Because of heterogeneity in the participation rates by age at invitation, we also produced age-specific self-selection factors by applying the following formula of our own derivation:

$$Sf_{age} = Sf\left(\frac{1-p}{1-p_{age}}\right) + Tf\left[1-\left(\frac{1-p}{1-p_{age}}\right)\right]$$

Here, p is the overall participation rate, 0.71; p_{age} is the age-specific participation rate; Sf is the overall odds ratio of those not participating as compared to those not invited; Tf is the odds ratio of those who would participate if invited when compared to those not invited, which cannot be observed directly but can be derived from the formula $(1 - p) \times Sf + p \times Tf = 1$; and, finally, Sf_{age} is the age-specific self-selection factor. The calculation of Sf_{age} rests on the assumption that the higher risk observed in non-participants is due to a small group of high-risk women who do not participate at any age. The high risk of these women is assumed to be due to lifestyle risk factors such as a large number of sexual partners, smoking, low use of health care services, and low socioeconomic status, all of which may also correlate negatively with screening attendance. Whenever the participation rate is lower, more women at low risk also fail to participate, thereby diluting the high risk of the non-participants and causing the self-selection factor to decline toward 1. When the participation rate is higher, the higher risk of non-participants is concentrated and the self-selection factor will become larger.

9 RESULTS

9.1 REGISTER QUALITY (I)

In 1998 to 2007 there were 1.9 million screening tests, 16,353 referrals for histological confirmation, and 6,646 precancerous lesions or cancers associated with the cervix registered in the screening programme. The coverage of screen-detected cervical lesion in the screening register was very high at all levels of lesion severity (see Table 7). The sensitivity of CIN diagnoses was also high in comparison with the other registers. However, the sensitivity of the ICC diagnosis was 69%, the lowest value of the three registers under comparison. The PPV of an ICC diagnosis in the screening register was found to be 77%. However, all cases were confirmed by FCR records and notifications as at least CIN2 neoplasias. The kappa value for inter-rater agreement between the MSR and the FCR was evaluated at 0.79.

Table 7: Sensitivity and coverage of register diagnosis by threshold of lesion severity

	ICC	CIN3+	CIN2+	CIN1+
Reference standard (*n*)	207	2,874	5,176	7,183
MSR				
Sensitivity (%)	68.6	81.3	89.2	92.5
Coverage (%)	100.0	99.4	99.3	99.4
HDR				
Sensitivity (%)	80.7	74.9	78.3	75.8
Coverage (%)	100.0	92.7	92.5	90.5
FCR				
Sensitivity (%)	98.6	79.5	NA	NA
Coverage (%)	99.0	79.5	NA	NA

9.2 FALSE NEGATIVES (II AND III)

The study material for studies II and III, which included smears taken by seven cytology laboratories during 1990–1999, covered 1,312,139 invitations and 953,610 smears. There were 9,062 (1.0%) referrals for colposcopy based on these smears

and 1,152 screen-detected CIN3+ cases. Audit cases (i.e., cases of CIN3+ preceded by Pap I or Pap II smears without referral for colposcopy) numbered 474. Of these, 395 (83%) were retrieved for review. The overall falsenegative rate was 38% among cases and 3.2% among controls (see Table 8). There were no significant differences in false-negative rates between invasive audit cases and CIN3/AIS audit cases. As expected, the falsenegative rates were higher among those with smears originally classified as Pap II (63%) than those originally classified as Pap I (24%). Those Pap I smears that were classified as at least borderline abnormal (ASC-US+) at review amounted to 53% of all audit cases and 14% of all controls.

Table 8: Cumulative frequencies of final cytology review result by smear status and original cytology

		Final review result		
		ASCUS+	LSIL+	HSIL+
		n (%)	*n* (%)	*n* (%)
Smears without referral in study	944,548			
Identified audit smears	474			
Audit smears in study	395	221 (56)	151 (38)	94 (24)
Cx cancer	58	31 (53)	22 (38)	15 (26)
Pap I	42	15 (36)	11 (26)	8 (19)
Pap II	16	16 (100)	11 (69)	7 (44)
CIN3/AIS	337	190 (56)	129 (38)	79 (23)
Pap I	210	87 (41)	49 (23)	29 (14)
Pap II	127	103 (81)	80 (63)	50 (39)
Controls in study	787	109 (14)	25 (3.2)	9 (1.1)
Pap I	743	83 (11)	14 (1.9)	5 (0.7)
Pap II	44	26 (59)	11 (25)	4 (9.1)

9.3 SCREENING-TEST VALIDITY (II)

Sensitivity and specificity of cytology in the screening laboratories in the review phase was evaluated with respect to the final review results (see Table 9). The sensitivity of the screening laboratory's review at the threshold of LSIL+ (equivalent to abnormalities usually requiring colposcopy) was 56%, and the specificity was 99%. Some 20% of these smears were deemed normal by the screening laboratories. At a threshold of ASC-US+, the sensitivity was 64% and specificity 97%.

Table 9: Sensitivity and specificity of the review performed by the screening laboratories with respect to final cytology results

	Final cytology (%)		
	ASC-US+	LSIL+	HSIL+
Sensitivity			
ASC-US+	64.2	79.6	85.4
LSIL+	32.7	55.7	66.2
HSIL+	18.8	31.8	47.6
Specificity			
ASC-US+	97.4	90.7	86.5
LSIL+	99.9	98.9	96.2
HSIL+	100.0	99.4	98.8

The inter-observer reproducibility of cytology was evaluated via comparison of the review results of the original screening laboratories with the review results of the reference laboratory (see Table 10). For the detail-level Bethesda categories, there was 69% agreement between the two reviewing laboratories. The unweighted kappa statistic indicated fair agreement, at 0.26. The results were also grouped according to the recommendation given. The three categories used were normal, intensified screening protocol for ASC-US and AGC-NOS, and LSIL or indication of more severe abnormality – considered a basis for referral. For these three categories, there was 84% agreement between the laboratories and a kappa value of 0.59, suggesting that the agreement was moderate. When these three recommendation categories were further analysed with a linear weighting for level of disagreement, the kappa statistic indicated substantial agreement (0.66).

Table 10: Inter-observer reproducibility of cytology

	Reference laboratory								
Screening laboratory	Normal	ASC-US	AGC-NOS	LSIL	ASC-H	HSIL	AGC-FN	Ca	Total
Normal	751	102	41	13	14	20	7	0	948
ASC-US	46	23	7	7	5	8	5	0	101
AGC-NOS	11	3	3	1	1	2	3	0	24
LSIL	4	2	0	8	0	7	3	0	24
ASC-H	10	3	2	2	3	2	1	0	23
HSIL	6	8	2	1	7	21	3	1	49
AGC-FN	0	1	3	0	1	3	2	0	10
Ca	0	0	0	1	0	1	0	1	3
Total	828	142	58	33	31	64	24	2	1,182

Cohen's unweighted κ statistic is 0.26. For the three categories of normal, intensified screening protocol, and referral, the unweighted κ = 0.59 and linearly weighted κ = 0.66.

9.4 PERFORMANCE INDICATORS (III)

Performance or process indicators for the seven laboratories participating in the cytology audit are presented in Table 11. All cross-sectional performance indicators were heterogeneous across laboratories. The referral rate varied from 0.3% to 1.2% (fourfold), with an average of 1.0%. The rate of recommendations for intensified screening varied from 2.8% to 10.2%, and the average was 5.4%. Consequently, the PPVs, especially for CIN3+ histology, showed considerable, eightfold variation, from 5% to 41%, with an average of 13%. False-negative proportions for the audited CIN3+ case smears at a threshold of LSIL+ ranged from 29% to 62% and had an average of 38%. Review sensitivity and specificity too were highly variable. However, there was less variation in the longitudinal outcome indicators. The audit case rate out of all smears taken by the laboratory varied by a factor of two, from 0.023 to 0.048, and the proportion of all CIN3+ cases that were screen-detected in the programme ranged from 69% to 85%. The two outcome indicators did not show statistically significant heterogeneity when formally tested by likelihood ratio tests.

Table 11: Performance parameters and outcome indicators by laboratory (%)

	Screening laboratory, with number of smears							
	A	B	C	D	E	F	G	All
	265,536	177,948	162,478	109,354	105,241	78,481	54,572	953,610
Attendance	68	73	76	69	81	79	74	73
Referrals	1.2	1.0	1.2	1.1	0.4	0.6	0.3	1.0
Intensified screening	6.9	3.2	3.2	8.3	2.8	10.2	4.4	5.4
CIN1/CIN3+ (ratio)	1.15	1.00	2.39	0.65	0.41	0.81	0.17	0.98
PPV (CIN3+)	13	12	5	16	31	18	41	13
PPV (CIN2+)	24	20	13	38	46	33	57	25
PPV (CIN1+)	38	32	24	48	58	48	63	37
False-negative rate*	30	29	44	62	32	50	39	38
Review sensitivity (LSIL+)	50.9	63.6	45.2	78.4	23.8	78.6	55.6	55.7
Review specificity (LSIL+)	99.5	96.7	100.0	94.6	100.0	100.0	97.8	98.9
Audit case rate†	0.030	0.028	0.025	0.048	0.044	0.027	0.023	0.034
Proportion detected by screen out of all CIN3+ cases‡	84	80	69	79	85	81	85	80

* Proportion of originally Pap I or Pap II smears in audit with final review cytology of LSIL+

† Audit CIN3+ cases diagnosed after Pap I or Pap II screen in 1996-1999 divided by all smears taken in the period

‡ In period 1996-1999.

9.5 AGE AT DIAGNOSIS AND MODE OF DETECTION (IV AND V)

Microinvasive carcinomas (FIGO stage IA) were most frequently diagnosed just after the age of 30 (see Figure 2). The frankly invasive cancer (FIGO stage IB+) frequency had two peaks, the first some years after age 35 and the other at around age 75. Cervical cancers that caused death were diagnosed predominantly toward older ages, and the frequency peaked at around 80. There were no deaths associated with cervical cancers diagnosed before age 25, and only nine cases with cancers were diagnosed before the age of 30.

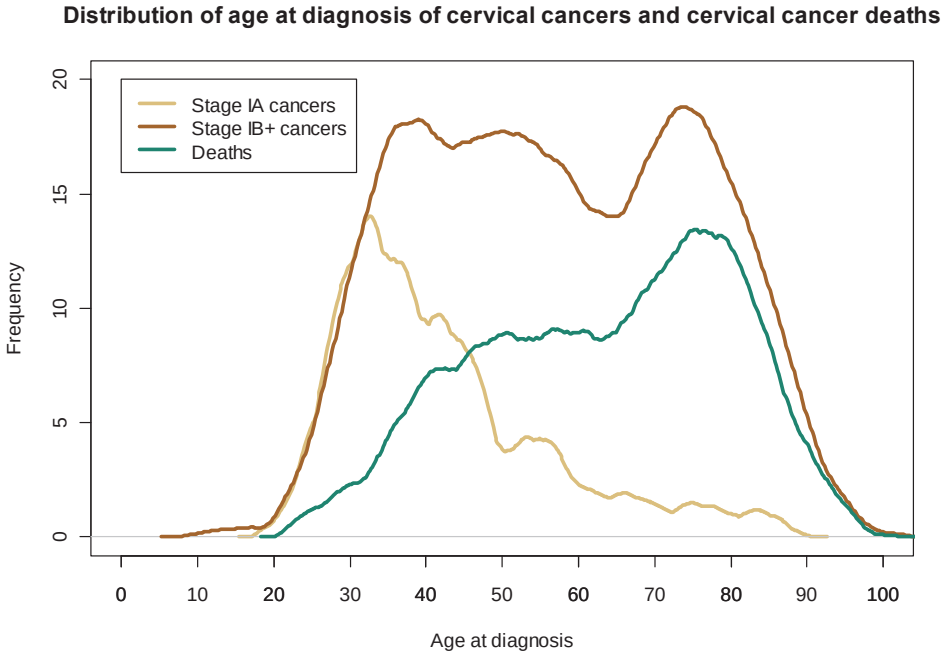

Figure 2. Age at diagnosis for all incident FIGO stage-IA and stage-IB+ cervical cancers in Finland in 2000–2009 (n = 1,548) and age at diagnosis for all cases of death attributed to cervical cancer in Finland in 2000–2009 (n = 545). Frequencies are smoothed by kernel density estimation for better visualisation.

Nearly one third of all cancers and more than half of the cancers leading to death were diagnosed more than five years after the last programme invitation (see Figure 3). The second largest group was that of the non-attenders; 28% of cancers and 24% of deaths occurred in this group. The screendetected cancers accounted for 11% of cancers but only 3% of deaths. Interval cancers were also less common as a

cause of death (13% of all deaths) than among all cancer diagnoses (19%). A small proportion (4.4%) of cancers was diagnosed before first programme invitation, and only 1.6% of the deaths were due to these cancers.

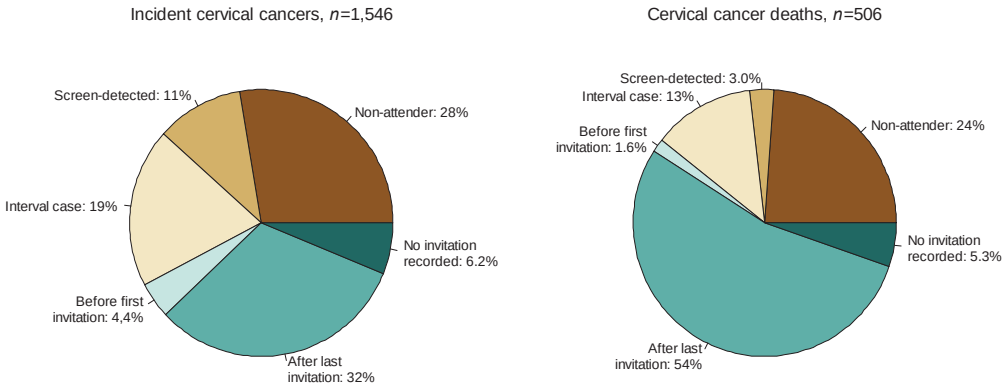

Figure 3. Cervical cancers and cervical cancer deaths (Finland, 2000–2009), that were included in the study, by mode of detection in relation to screening.

9.6 EVALUATION OF SCREENING EFFECT

The proportions of those without invitation in the index round were similar between cases and controls for both the incident cervical cancers and the deaths (Table 12; papers IV and V). This was expected, as lacking an invitation was mainly due to diagnosis outside the programme target age range so indirectly matched for. However, clear differences were observed in the proportions of non-attenders and attenders. Out of the incident cases, 32% were non-attenders in the index round, as compared to 20% of their controls. The difference was even larger among cancer deaths, for which invited non-attenders were more than twice as common as among the controls. Four cases of cancer were diagnosed after a negative HPV test, whereas there were no deaths due to cancers in this category. No deaths were due to cancers diagnosed after a colposcopy in the index round.

Table 12: Screening status of index round at time of cancer diagnosis for cases and on corresponding date for matched controls

	Incident cervical cancers				Cervical cancer deaths			
	Cases		Controls		Cases		Controls	
	n	%	n	%	n	%	n	%
Total	1,546	*100.0*	9,276	*100.0*	506	*100.0*	3,036	*100.0*
No invitation	686	*44.4*	3,994	*43.1*	308	*60.9*	1,818	*59.9*
Before first invitation	91	*5.9*	503	*5.4*	8	*1.6*	36	*1.2*
Over screening age*	487	*31.5*	2,920	*31.5*	270	*53.4*	1,609	*53.0*
Other reason	108	*7.0*	571	*6.2*	30	*5.9*	173	*5.7*
Non-attender	494	*32.0*	1,813	*19.5*	127	*25.1*	342	*11.3*
Never attended†	414	*26.8*	1,595	*17.2*	98	*19.4*	201	*6.6*
Lapsed attender	80	*5.2*	218	*2.4*	29	*5.7*	141	*4.6*
Screened	366	*23.7*	3,469	*37.4*	71	*14.0*	876	*28.9*
Negative cytology	254	*16.4*	3,190	*34.4*	64	*12.6*	817	*26.9*
Borderline cytology	94	*6.1*	164	*1.8*	6	*1.2*	37	*1.2*
Negative HPV test	4	*0.3*	96	*1.0*	0	*0.0*	15	*0.5*
Referral non-compliance	1	*0.1*	1	*0.0*	1	*0.2*	1	*0.0*
Negative histology	6	*0.4*	9	*0.1*	0	*0.0*	2	*0.1*
Positive histology	7	*0.5*	9	*0.1*	0	*0.0*	4	*0.1*

* Diagnosis five years or more after last programme invitation

† During period 1990-2009.

9.6.1 SCREENING EFFECTIVENESS (IV AND V)

Self-selection-corrected odds ratios of the association between cervical cancer and participation in programme screening indicate that there was little or no effect of screening at ages 25–29 (see Figure 4). Programme participation between ages 30 and 39 was associated with a 21–27% reduction in cancer risk, but this effect estimate was not significant. From the age of 40, the risk reduction exceeded 50% for all five-year age bands, but the point estimate for the last five-year age group, 65–69, was non-significant.

Age-specific odds ratios of programme participation by morphology and for cervical cancer death were estimated in 15-year age groups for more robust estimates (see Table 13). The ORs for SCC were similar to those for all cancers. However, there was a suggestion of stronger effect in the older age groups, especially with death as an outcome. The effect on the risk of adenocarcinoma was smaller and the estimates non-significant for death due to adenocarcinoma in the various age groups and overall. Point estimates suggested a higher impact of screening on the risk of death than on the risk of cervical cancer, except for adenocarcinoma. The overall reduction of cervical cancer risk produced by one age-group-programme test was estimated at 47% (95% CI: 38–54%) whereas overall reduction in risk of cervical cancer death was estimated at 66% (95% CI: 51–86%).

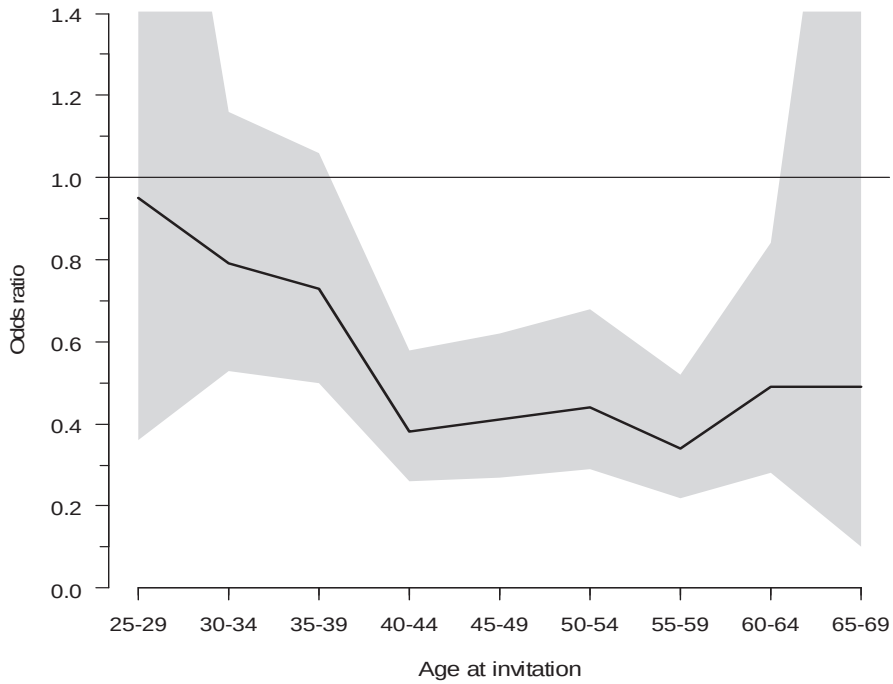

Figure 4. Odds ratios of the association of screening programme participation and cervical cancer in the following five-year period. Estimates are corrected for self-selection bias, with self-selection factor 1.29 and a participation rate of 71%. Confidence intervals (95%) are shaded grey.

Table 13: Odds ratios for the association between cervical cancer or cervical cancer death and screening participation by cancer morphology

		Cervical cancer		Cervical cancer death	
Cancer morphology	Age at invitation	Crude OR	Corrected OR*	Crude OR	Corrected OR*
All cervical cancers	All (25–69)	0.37	0.53 (0.46–0.62)	0.19	0.34 (0.14–0.49)
	25–39	0.56	0.81 (0.63–1.05)	0.44	0.70 (0.33–1.48)
	40–54	0.30	0.44 (0.35–0.56)	0.18	0.33 (0.20–0.56)
	55–69	0.25	0.37 (0.27–0.52)	0.15	0.29 (0.16–0.54)
Squamous carcinoma	All (25–69)	0.34	0.50 (0.41–0.61)	0.12	0.22 (0.13–0.36)
	25–39	0.62	0.91 (0.67–1.24)	0.61	0.97 (0.39–2.41)
	40–54	0.27	0.40 (0.29–0.54)	0.06	0.11 (0.05–0.28)
	55–69	0.16	0.23 (0.14–0.37)	0.09	0.17 (0.07–0.44)
Adenocarcinoma	All (25–69)	0.47	0.69 (0.53–0.91)	0.42	0.75 (0.43–1.31)
	25–39	0.41	0.59 (0.36–0.98)	0.32	0.50 (0.09–2.76)
	40–54	0.42	0.61 (0.41–0.91)	0.65	1.21 (0.53–2.75)
	55–69	0.60	0.88 (0.50–1.54)	0.28	0.55 (0.23–1.35)

* Overall ORs were corrected via participation rate (*p)* 0.71, self-selection factor (*Sf)* 1.29 for incidence and 1.45 for death. Age-group-specific self-selection factors were used for ORs associated with death: *Sf* 1.30 and *p* 0.62 for the age group 25–39, *Sf* 1.51 and *p* 0.74 for the age group 40–54, and *Sf* 1.63 and *p* 0.77 for the age group 55–69. The overall values were used for all ORs associated with incidence. 95% CIs of estimates are indicated.

Small adjustments in the point estimates were observed when age-specific correction was used also for the ORs of the association between screening participation and cancer incidence (see Table 14, presenting results not previously published).

Table 14: ORs for the association between screening participation and cervical cancer, with overall or age-group-specific corrections for selection bias

Age at invitation	OR with overall correction	Participation rate	Self-selection factor	OR with age-specific correction
Overall	0.53 (0.46–0.62)	0.71	1.29	NA
25–39	0.81 (0.63–1.05)	0.62	1.19	0.75 (0.58–0.98)
25–29	0.95 (0.36–2.49)	0.56	1.15	0.85 (0.37–1.93)
30–34	0.79 (0.53–1.16)	0.60	1.18	0.72 (0.49–1.07)
35–39	0.73 (0.50–1.06)	0.66	1.23	0.70 (0.48–1.01)
40–54	0.44 (0.35–0.56)	0.74	1.33	0.46 (0.36–0.58)
40–44	0.38 (0.26–0.58)	0.72	1.30	0.39 (0.26–0.58)
45–49	0.41 (0.27–0.62)	0.73	1.33	0.43 (0.28–0.64)
50–54	0.44 (0.29–0.68)	0.76	1.38	0.47 (0.31–0.73)
55–69	0.37 (0.27–0.52)	0.77	1.40	0.41 (0.29–0.57)
55–59	0.34 (0.22–0.52)	0.77	1.40	0.37 (0.24–0.57)
60–64	0.49 (0.28–0.84)	0.78	1.42	0.54 (0.31–0.93)
65–69	0.49 (0.10–2.41)	0.75	1.35	0.53 (0.17–2.62)

9.6.2 DURATION OF SCREENING EFFECT

We explored the duration of any risk reduction associated with programme screening by defining exposure as participation in the round preceding the index invitation among those without a screening test at index (see Table 15; papers IV and V). Overall OR estimates were significant, 0.76 (0.59–0.99) for cervical cancer and 0.48 (0.28–0.84) for death. Again, effect seemed to differ across age groups. A smear at the age of 55–69 was associated with a significantly reduced risk of both cancer incidence and cancer death, whereas at ages 25–39 and 40–54, ORs were non-significant.

Table 15: Odds ratios for the association between cervical cancer or cervical cancer death and screening participation in the five-yearly invitation before index

	Cervical cancer	üüüüüüüüüüü
Age at invitation	Corrected OR (95% CI)*	Corrected OR (95% CI)*
All (25–69)	0.76 (0.59–0.99)	0.48 (0.28–0.84)
25–39	1.14 (0.73–1.78)	1.73 (0.44–6.80)
40–54	0.84 (0.56–1.26)	0.58 (0.28–1.23)
55–69	0.35 (0.20–0.62)	0.18 (0.05–0.62)

* Corrected with participation rate 0.71 and self-selection factor 1.29 for incidence and 1.45 for death. Age-group specific ORs with death as an outcome were corrected by means of age-group-specific values.

9.6.3 DURATION OF LOWERED RISK AFTER A NEGATIVE SCREENING TEST

The risk of cancer was analysed in overlapping one-year periods up to six years after a negative screening test (see Figure 5, presenting unpublished results). The OR remained below 1 for the whole period overall but developed differently in the different age groups. After 3.5 years, the OR was close to 1 for those screened as negative below the age of 40, whereas there was a more persistent reduction of risk after negative smears at 40–54 and at 55 and over.

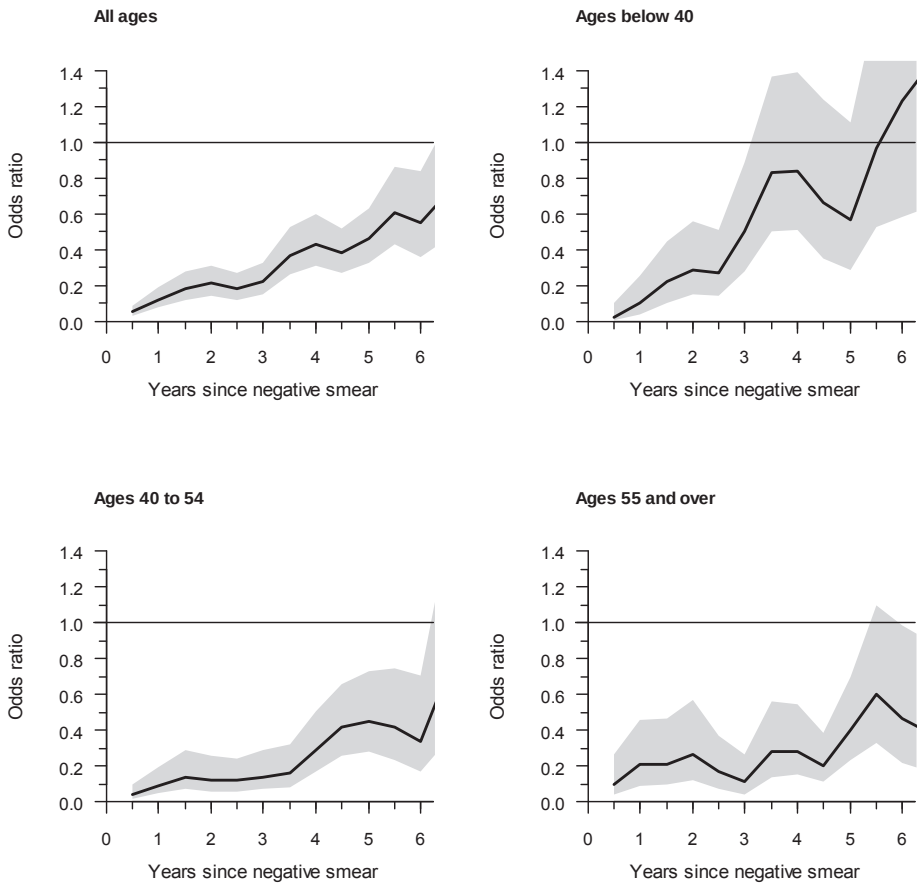

Figure 5. Odds ratios of cervical cancer after a negative programme smear as compared to no such test since 1990, by age group, estimated with conditional logistic regression in overlapping one-year periods, with 95% confidence intervals shaded grey.

9.6.4 EFFECT OF THE LAST PROGRAMME SCREEN (IV AND V)

We estimated the OR associated with cervical cancer or death with diagnosis at any age over 65 and screening participation among those invited at age 65. For cervical cancer, the corrected OR for this association was 0.66 (0.29–1.50), and for death the corrected OR was 0.25 (0.03–2.20). In order to capture more cases in the analysis, we also estimated ORs for cervical cancer or cancer death associated with a last programme screen at the age of 55, 60, or 65, regardless of invitation status (see Table 16). Significant reductions in the risk of both cervical cancer and cervical cancer death with diagnosis at ages 66–80 were found with a last screen

at 60 and 65. There was a tendency for greater effect with a last test at a later age, particularly when the time of diagnosis was restricted to ages 71–80.

Table 16: ORs for association of cervical cancer and last programme screen

	Age at last screen	OR (95% CI) Cervical cancer	Cervical cancer death
Diagnosis at age 66–80	none*	reference	reference
	55	0.54 (0.26–1.14)	0.45 (0.09–2.17)
	60	0.49 (0.33–0.73)	0.41 (0.21–0.78)
	65	0.41 (0.23–0.74)	0.38 (0.16–0.90)
Diagnosis at age 71–80	none*	reference	reference
	55	0.77 (0.14–4.22)	NA†
	60	0.47 (0.26–0.88)	0.43 (0.15–1.23)
	65	0.23 (0.08–0.63)	0.24 (0.06–1.00)

* No age-group programme screen at 55 or later registered in the study period

† No cases.

10 DISCUSSION

10.1 REGISTER VALIDATION

The screening register's coverage of cervical dysplasias and cancers was found to be excellent; hence, the information contained in the register is suitable for monitoring activities. The grade of the lesion registered in the MSR was usually based on the first pathology report and therefore not always the same as the final diagnosis as registered in the FCR. For this reason, the sensitivity of the ICC diagnosis was only 69% even though all of these cases were recorded with some histological diagnosis. Of the comparison registers, the HDR had high coverage of cancers but lower coverage of dysplasias. There is a development of registering more types and an increasing proportion of health-care contacts in the HDR, and the coverage of the lower-grade dysplasias can be expected to improve further with time. The FCR receives active reports from many sources. Hence, especially for cancer, both coverage and accuracy are very high. The coverage of lesions with morphological codes included in the CIN3 category was found to be somewhat lower than that for cancer and than in the other registers. This is a result of both lingering underreporting to the FCR of these precancer lesions by clinicians and pathologists and, in this study context, also possible over-reporting in the registers contributing to the reference pool of diagnoses (MSR and HDR). Therefore, the real coverage of CIN3+ lesions in the FCR is probably higher than the 80% reported here, if all of these diagnoses should be verified by extraction from the original patient records in hospital archives. CIN1 and CIN2 lesions are not registered in the FCR.

If specific individual-level diagnoses are needed, especially with respect to the screen-detected cervical cancers, the data on MSR sensitivity (69%) and PPV (77%) suggest that data retrieval through linkage of health-care registers should be considered for accurate diagnoses.

There are no previous studies of the validity of the screening register data on histological diagnoses. The validity of the HDR data for cervical disease has not been examined, but there have been publications on the accuracy of cardiovascular diagnoses. Sensitivity for stroke and coronary heart disease in this register has been reported as 83–85% and the PPVs for these diagnoses as 83–92% (Pajunen et al. 2005, Tolonen et al. 2007). The sensitivity for ICC in our study was 81%, with a PPV of ICC diagnosis of 83%, roughly in line with the findings cited for cardiovascular diagnoses.

The FCR is regarded as nearly exhaustive with regard to cancer diagnoses overall, thanks to the use of multiple notification sources (Teppo et al. 1994). In our study, which covered 10 years and 207 cases, there were only three cases of ICC missing from the original FCR records. One of these was originally registered as an unspecified female-genital-organ cancer. This translates to high coverage and sensitivity, both 99%, also specifically for the ICC diagnosis.

10.2 CYTOLOGY AUDIT AND TEST VALIDITY

For the screening service, the relevance of the cross-sectional validity measures presented in Table 9 may be somewhat limited, because they are derived from comparison of the ratings of two observers probably using variable criteria to call cytological smears. The validity estimates are, therefore, a composite of true sensitivity and specificity, precision or inter-observer reproducibility, and also variable criteria used by the service screeners, reference laboratory, and expert panel in the cytodiagnosis. Here, low sensitivity indicates a difference between screening-service cytology and expert-panel cytology. Practical conclusions may be difficult to draw. We saw that, while a larger proportion of audit cases were abnormal according to final cytology, the differences in cytodiagnostic criteria also resulted in 3% of the controls being classified as needing a referral for colposcopy. If this were generalised to routine screening, this would mean a tripling of the colposcopies performed in the programme, where referral rates are currently around 1%. This may not be justifiable in light of the benefits on offer in terms of disease prevented. The study illustrates clearly the importance of including controls in any cytology audit. Because of the controls, it was evident that any increase in sensitivity of the cytodiagnosis requires that some of the specificity of the screening test be sacrificed.

Another, longitudinal, measure of sensitivity could be the proportion of CIN3+ screen-detected out of the total number of CIN3+ cases arising in the study period among those screened. This proportion amounted to 71% overall (1152/(1152-474)). Similarly, specificity can be defined as the proportion of those without CIN3+ in the study period who were not referred for colposcopy after their programme smear. This proportion was 99.2% ((944548-474)/(953610-474-1152)).

The reported sensitivity and specificity of cytology vary widely by setting. This is partly a function of the thresholds of both cytology (test threshold) and the gold standard (disease threshold). The gold, or reference, standard is usually histology but can also be expert-panel cytology or longitudinal disease outcomes. In addition, the inherent subjectivity of cytology as a test and also of a gold standard that relies on colposcopy and histology (both subjective) will introduce variability in the validity estimates observed in different studies. The specificity will further

suffer in particular from verification bias if only the test positives are assessed for the presence of disease. This is the case in service screening and also often in studies estimating the validity of cytology. If complete histological verification is used, the observed sensitivity will be lower and the specificity higher. In a meta-analysis, the magnitude of the effect of verification bias was demonstrated by a decline in sensitivity when complete verification was used, from 0.67 to 0.52, and an increase in specificity, from 0.73 to 0.96 at a test threshold of LSIL+ and a disease threshold of CIN1+ (McCrory et al. 1999). The corresponding changes at disease threshold CIN2+ were 0.83 to 0.77 for sensitivity and 0.61 to 0.92 for specificity.

The variability in the validity of cytology was demonstrated by the ranges of sensitivity and specificity cited in a review of test-accuracy studies (Nanda et al. 2000). For those studies that used low-risk women presumably more representative of the general screening population and complete- or random-sample verification of negative test results, the reported sensitivity still ranged from 30% to 87% and the specificity from 86% to 100% for a test threshold of LSIL+ and a disease threshold of CIN1+. For disease threshold CIN2-3+, the ranges were 44% to 99% for sensitivity and 91% to 98% for specificity. Without these eligibility criteria in the studies reviewed, the variability was even more pronounced. The sensitivity results of the original screening-laboratory review with threshold LSIL+ in our cytology audit (i.e., 50% when the original six laboratories were used, in Paper II, or 56% when calculated from all seven laboratories participating in the audit – see Table 9; papers II and III) are best compared to an LSIL+/CIN1 threshold combination if histology is used as the reference standard. These values sit comfortably in the middle of the range of 30–87% reported by Nanda and colleagues. However, the corresponding specificity value of 99% of the Finnish screening-laboratory cytodiagnosis is at the higher end of the range 86–100% reported by Nanda et al. This observation is supported by the relatively low referral rates in the Finnish screening programme when compared to other European programmes, this despite demonstrably high effectiveness in cancer prevention (Ronco et al. 2009).

The choice of outcome in the cytology audit was CIN3+, which is a proxy for the invasive cervical cancer outcome that screening is to prevent. There is a limitation in that not all of the CIN3/AIS lesions would progress to cancer and, so, the number of audited 'failures' may be too high. However, there are some advantages to using CIN3+ instead of cancer as outcome in the audit. If there is a quality problem, it is desirable to know of this as early as possible, so that timely feedback can be produced for the service provider. By using CIN3+ as outcome in the evaluation, one can get audit results from larger numbers or earlier than would be possible with invasive cancer as audit outcome. Also, the results suggested that the false-

negative proportions are very similar among the audited CIN3/AIS and cancer cases. It would seem that this outcome can safely be used if the follow-up time is appropriately chosen.

The false-negative rate observed in our audit was 35% (of archived slides of subsequent cervical cancer or interval CIN3+ cases), as reported in Paper II with data from six laboratories and 38% when the whole set of seven laboratories was used (see Table 8). Only one cytology audit was found in the literature review that included smears reported originally as both Pap I and Pap II without referral; there, the proportion of smears reviewed as clearly abnormal was 53% (Kenter et al. 1996). Most studies have included only clearly negative (i.e., Pap I) audit smears in the review. These studies, with ICC as outcome, observed clearly positive false-negative proportions, 19–31% (as shown in Table 3). The corresponding Pap I/LSIL+ result in our study was 24% ((11+49)/(42+210)) (see Table 8), which is in line with the values found in other programmes. By a third measure of false negativity, where negative or Pap I smears are reviewed for any abnormality, including borderline changes, the range of proportions found in previous publications was 21–71% (see Table 3). The range is wide probably in part because of variable ways of defining abnormal smears, as some studies may have included non-neoplastic abnormalities whereas others may have included only cytological abnormalities indicative of dysplasia. Our estimate with application of these thresholds (Pap I/ASC-US+) was 40%, well in line with these previously reported values.

On the assumption that the false-negative rates observed in the review material were representative of the whole audit-case population, the analytical failures of the screening test accounted for only 11% of the total CIN3+ burden in the screened population.

As the cross-sectional sensitivity and also longitudinal clinical sensitivity and negative predictive value of HPV tests are favourable in comparison to cytology, a switch to HPV-based primary screening protocols would likely further reduce false-negative rates (Kitchener et al. 2011). The large variations in cytology tests' validity between settings that arises from the subjective nature of the analysis would also be alleviated by automated and objective HPV analysis. The challenge of using HPV tests in primary screening is the handling of lower test specificity for cervical lesions that require treatment.

10.3 SCREENING PERFORMANCE INDICATORS

There were large differences in all cross-sectional performance indicators among the seven screening laboratories in the cytology audit. In general, the differences can be interpreted as representing laboratory policies of reporting cytology with

differential relative emphasis on the sensitivity vs. specificity of the cytodiagnosis. It is also possible that risk factors, mainly HPV prevalence, have been variable across the regions covered by the laboratories (Lehtinen et al. 2006). The CIN detection profile, which can be described by the ratio between CIN1 and CIN3+ detected by screening, reflects the reporting policy of the laboratory. The CIN1/CIN3+ ratio ranged from 0.2 to 2.4, with an average of 1.0. Specificity-oriented laboratories have a low CIN1/CIN3+ ratio, low referral rate, high test specificity, and high PPV (laboratories E and G), while the opposite is true for laboratories that prioritise sensitivity (laboratories A, B, and C) (Paper III). In theory, operating a screening programme with high sensitivity should be more effective at removing precancerous lesions from the population and hence preventing cervical cancer incidence and mortality. Orientation toward high specificity, on the other hand, should limit the adverse effects of screening such as the psychological burden of positive test results and number of colposcopies, biopsies, and possibly also conisations. However, population-based effects are strongly influenced also by screening uptake by women at risk and the quality of follow-up and management of screen positives.

The audit-case rate among screened women varied to a much lesser degree than the cross-sectional indicators did. No significant heterogeneity was observed in this measure upon formal testing.

Considerable differences in performance indicators have been reported between screening centres within individual countries (Kotaniemi-Talonen et al. 2007, Blanks 2008). When one compares indicators from national programmes, the differences are even greater (Ronco et al. 2009). For example, the aggregate data for Italian screening programmes in 2008 indicated an overall referral rate of 2.4% with a PPV of referral for CIN2+ lesions of 16% (Ronco et al. 2010b), compared to the referral rate of 1.0% and PPV for CIN2+ of 25% observed in our study overall (Paper III). The detection rates of these two programmes differ less from each other, 3.1/1000 screened women in Italy compared to 2.4/1000 in Finland. In the Netherlands, a 1.4% referral rate produced a PPV of 42% for CIN2+ lesions and a detection rate for CIN2+ of 6.4/1000. Corresponding values for the English programme were a referral rate of 2.7%, PPV of 49%, and CIN2+ detection rate of 11/1000 (Ronco et al. 2009). It is difficult to draw conclusions from these comparisons. Not only do the values depend on the criteria used for the cytodiagnosis, but also the disease prevalence, thresholds for intensified screening recommendations and referral for colposcopy, quality of histological verification, compliance rates, and magnitude of any opportunistic screening activity will have an impact. Significant opportunistic screening occurs in Italy and Finland (Ronco et al. 2010a, THL 2011), whilst this activity is now far less commonplace in the Netherlands, because of withdrawal of subsidies. Opportunistic screening may also affect the number of potential false-negative audit cases observed in a population by detecting cases of severe dysplasia

that would have regressed before the next scheduled programmatic screening event. Lifetime regression rates of CIN3 or equivalent lesions have been estimated in cohort studies to be between 12% and 32% (Syrjänen 2009); in the five-year interval between programme invitations, the proportion should be much less.

There are no recommended target values for performance indicators in cervical screening as there are for breast cancer screening. Because of the complexity of the screening process, from identification and composition of the target population and its disease and risk factors' prevalence to the quality of diagnostic confirmation, therapy, and follow-up, the relevance of any recommendations may be restricted to individual programmes or regions. This difficulty applies to the optimal values of referral rates, CIN detection profiles, and related indicators while it is clear that the attendance rate in the invitational screening programme and compliance with referral recommendations and follow-up should be as high as possible.

Even though the reporting policies were found to vary to a large degree between laboratories in the screening programme, no significant variation was observed in the rates of missed progressive lesions. It is possible that the differences in test validity and detection rates for progressive lesions are smaller than for non-progressive lesions. It is also possible that intensified programme screening and opportunistic screening in the interval between age-group invitations catch missed lesions before they have a chance to progress to invasive or pre-invasive disease. However, it is important to continue to produce regular feedback by using outcome information from follow-up after screening in order to harmonise reporting and recommendation policies and to avoid unnecessary adverse psychological and physical effects of the follow-up and management of smears reported as abnormal.

10.4 AUDIT OF SCREENING HISTORY

10.4.1 MODE OF DETECTION

Screen-detected cancers have a favourable prognosis when compared to symptomatic cancers (van der Aa et al. 2008, Andrae et al. 2012), because diagnosis in the preclinical phase allows more successful and often less aggressive radical and curative treatment even after adjustment for stage at presentation. On the other hand, screen-detected cancers should, as should all other cancers, be considered failures of the screening programme, which primarily aims to identify and treat precancerous lesions in order to prevent the development of invasive disease. Stage at diagnosis is a separate but related dimension. Screen-detected cancers (non-

symptomatic cancers) are more often of lower stage than symptomatic cancers, and the microinvasive stage is nearly always screen-detected. Cancers diagnosed at the microinvasive stage very rarely proceed to cause death in settings with adequate health-care capabilities (Paper V). Therefore, some experts consider only frankly invasive (stage IB+) cancers to be failures of the screening programme.

Screen-detected cancers constituted 11% of the total cancer burden in Finland in 2000–2009. Van der Aa et al. (2008) reported that 35% of the cancers diagnosed in the Netherlands were screen-detected in the age groups targeted by screening in 1992–2001. The equivalent proportion in our setting was 19% when the denominator is restricted to those invited to take part in the programme (Paper IV), still far below the Dutch figure. The extent of opportunistic screening hampers the comparison, as the distinction of screen-detected and symptomatic cancers only holds when all screening smears are in the programme. Certainly, non-symptomatic cancers are diagnosed outside the Finnish programme, as indicated by the many microinvasive carcinomas found among nonattenders and during the between-screening intervals (Paper IV). The Dutch study did not report microinvasive carcinomas separately. High proportions of screen-detection have been reported from the UK. Of 133 cancers diagnosed in two London boroughs in 1999–2007, 49% were screen-detected (Herbert et al. 2010), and from among 382 cancers diagnosed in Southampton in 1985–1996, 33% were screen-detected (Herbert et al. 2009a). In the Swedish audit, 25% of all cancers (Andrae et al. 2008), and 32% of cancers at screening ages (Andrae et al. 2012) diagnosed in 1999–2001 were screen-detected. In addition to opportunistic screening, the definition applied for screen-detection may account for part of the difference observed. We have used the presence of a referral for colposcopy within 12 months before diagnosis as evidence of screen-detection, as recommended in the European guidelines (Arbyn et al. 2008a). In the Netherlands, the pathology and cytology register PALGA can provide this information (Casparie et al. 2007). The Swedish audit defined screen-detection as a (presumably abnormal) smear 1–6 months before diagnosis, and the UK audits have used clinical criteria as determinants of screen-detection (Herbert et al. 2009b).

No information was available for comparison with respect to deaths due to screen-detected cancers as a proportion of all deaths due to cervical cancer. At 3%, this proportion was much lower than for all cervical cancers in our material.

10.4.2 PREVENTIVE SCREENING STATUS

In order to achieve high effects of screening on incidence and mortality, coverage and attendance are a priority in all screening programmes and usually also constitute the most important area for improvement in the effectiveness of established

programmes. Coverage by invitation is high in the Finnish programme. The current coverage is almost complete: 98–100% of those at the recommended target ages in 2009 (see Table 5), but historically there has been slightly more variation, with 6–7% of women belonging to the currently recommended age groups lacking a registered invitation in the audit study period, 1990–2009. However, this figure is an overestimate, because it includes women registered as residents in the interval between index invitation and date of diagnosis. In addition, invitations have been under-reported to the centralised mass screening registry in a small number of cases, inflating the figure further. The proportion of cancers diagnosed in women after the last invitation was 32%, and more than half of the deaths were attributed to these cancers. In other well-established programmes, the equivalent proportion was similar or slightly lower. In the Swedish audit, 25% of cancers were diagnosed at ages above 65, corresponding to time after the last screening round (Andrae et al. 2008). In the Southampton audit, 30% of cancers were diagnosed in this age group (Herbert et al. 2009a). The proportion can be expected to increase as screening removes cancers at the target ages.

The attendance rate was 71% overall in the study period, with a very slow deterioration observable over the past decade, mainly in ages under 40. The second largest group by preventive screening status at index round among both cancers and cancer deaths was that of the non-attenders. This group accounted for 32% of the incident cancers and 25% of the deaths, indicating that further efforts to improve attendance are needed. Those developing cancer despite attending screening are of special concern in the screening-service audit and, especially, for the feedback to screening laboratories. In the previously published screening-history audits listed in Table 2, the range of the proportion of cancers diagnosed after participation in screening in a comparable defined period of time was 19–54%. Our result of 24% is at the lower end of this range. For deaths in our study, the corresponding proportion was 14%. Within this category, we can further identify the potentially false-negative screening tests, which amounted to 16% of cancers and 13% of deaths, as compared to the range of 9–37% in previous studies (again, see Table 2).

There were only four cancers diagnosed after a negative HPV screen, 0.3% of all cancers, compared to 1.0% of controls. The high sensitivity and NPV of HPV testing is especially visible in the corresponding numbers among deaths: no deaths were observed due to cancers diagnosed after a negative HPV screening test, compared to 15, or 0.5%, of the controls having a negative HPV test at index screen. The difference in the proportion of negative HPV tests at index between the two control populations is due to the difference in the distribution of diagnosis over calendar time relative to the HPV trial. Diagnoses were made in 2000–2009 in the incidence data and 1990–2009 among the deaths, and the HPV DNA test was introduced in 2003.

Compliance with referral seems excellent; only one case of non-compliance was found among the cases and another among the controls. A similar observation was made previously for an earlier period of screening (Viikki et al. 2000). Management failures were relatively rare. Under 1% of all cancer diagnoses had a colposcopy at index screen. There were no deaths due to cancers diagnosed after a colposcopy in the index round, no matter the histological outcome of the visit. In comparison, a systematic review of screening-history audits reported an average of 12% of cancer cases as failures of follow-up after a truly abnormal smear prior to diagnosis. Individual studies reported proportions ranging from 3.2% to 50% (Spence et al. 2007).

These data on the screening history of cases justify further enquiry into the viability of extending screening invitations also to women aged 65, as 32% of cancers and 53% of deaths have diagnoses after this age. As expected, non-attenders constitute the other important focal point for improvement, with 32% of cancer cases and 25% of cancer deaths belonging to this category. Efforts to increase attendance are already under way. For instance, promising results have been achieved by sending self-sampling kits to women who fail to participate (Virtanen et al. 2011). Improving the sensitivity of the screening test can have a theoretical maximum effect of further reducing incidence by 17% and mortality by 13% (corresponding to the proportions of cases having a negative screening test at index).

10.5 EFFECTS OF PARTICIPATION IN SCREENING

10.5.1 EFFECTIVENESS OF ORGANISED SCREENING

The case–control analysis of screening exposure estimated the association of screening participation with any reduction in risk of either cervical cancer or death from cervical cancer with a diagnosis in the five-year screening interval up to and including any screen-detected cancers in the following screening event. Only participation in the organised screening programme counted toward screening exposure. It is likely that some of the women classified as non-exposed to programme screening had been opportunistically screened during the window of exposure (THL 2011). The risks of cervical cancer and cervical cancer death of programme non-attenders may therefore be somewhat lower than would be the case without opportunistic screening, leading, in turn, to lower estimates of the risk reduction associated with programme screening attendance. Bearing in mind that the estimates refer to programmatic screening attendance against a backdrop of opportunistic screening, we found that a single screening test in the programme reduced the risk of

cervical cancer of any morphological type or stage in the next five-year interval by 47% overall. The corresponding reduction in the risk of death with diagnosis in the same period was 66%. However, the effect was strongly age-dependent, with smaller reductions for participation before the age of 40 and uniformly large risk reductions associated with participation at age 40 and above. A tendency toward lower risk reduction with lower age was also observed within the age range 25–39.

Age-dependent effectiveness or efficacy of screening has recently been reported also from other screening programmes. A large audit from the UK reported highly age-dependent ORs for participation in screening. Screening at ages 22–24 had no effect with an OR over 1, but after this age the OR rapidly dropped such that at 30–37, screening participation was associated with a reduction in risk of 43–60% and at 40–64 the reduction in risk was 64–82% (Sasieni et al. 2009a). Two case–control studies from Italy have reported lower effects of screening in women under the age of 40. A study describing the programme in Trento found no association with screening participation for women under 40 for cancers of any morphological type (OR: 1.00, 95% CI: 0.18–5.65) or for squamouscell cancers (OR: 2.03, CI: 0.20–20.4), whereas the point estimate indicated a lowered risk of adenocarcinomas (OR: 0.25, 0.02–4.0) (Crocetti et al. 2007). The small numbers render these estimates unreliable, but they suggest a pattern similar to that found in our material. Data from Florence also suggested smaller effects for screening participation in the previous five years at ages under 40 although still significant and considerable (OR: 0.32, CI: 0.11–0.95) (Zappa et al. 2004). The ORs for the other age groups were 0.11 (0.04–0.33) for women aged 40–49, 0.08 (0.02–0.31) for women aged 50–59, and 0.22 (0.06–0.83) for women 60–69. Lower ORs of the association of any screening participation and cancer diagnosed under the age of 30 were observed in South Africa, as compared to diagnosis at ages 30 and above (OR: 0.7, CI: 0.3–2.1 versus OR: 0.3, CI: 0.2–0.4) (Hoffman et al. 2003). In the Swedish audit, equal effects were reported initially for screening participation across all age groups (Andrae et al. 2008). In a subsequent discussion between the authors of the Swedish and the UK audits, additional analyses were presented wherein the effect on IB+ cancers diagnosed at age 27–29 persisted but no effect could be demonstrated for women aged 23–26 (Andrae et al. 2009). No corrections for self-selection bias were made in any of the above studies.

The trends for cervical cancer in Finland show that the most dramatic declines in incidence since the start of the programme (with rollout during the 1960s) have been observed for ages 40–49 (from 42/100,000 in 1965 to 6/100,000 in 1990) and 50–59 (from 48/100,000 in 1965 to 7/100,000 in 1990) (see Figure 6). A smaller but still substantial decline can be seen for ages 30–39 (15/100,000 to 4/100,000), albeit from a lower level, but incidence was already falling in this age group when the screening programme was launched, in contrast with the 40–59

age group, for which incidence was increasing. Also, since 1990, incidence has again been increasing for ages below 40, while decreasing trends have continued in all other age groups. It seems likely that there has been some impact of screening on incidence among those below 40 also but that this effect is smaller than that in older women. The current case–control data support this view.

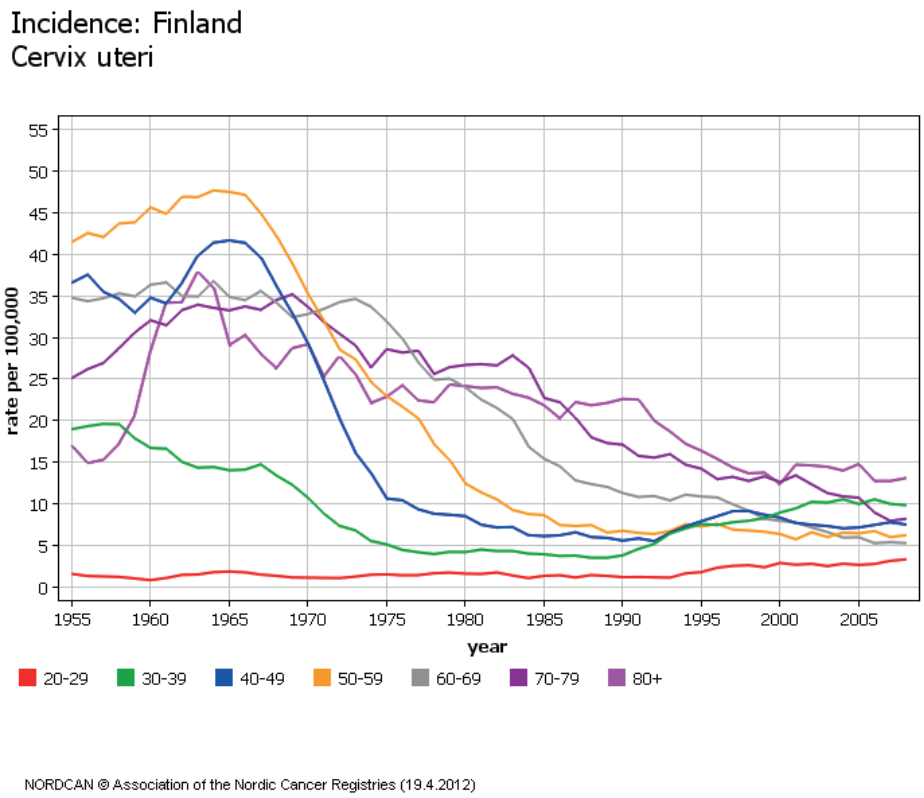

Figure 6.

The reasons for the weak association between screening attendance and cervical cancer risk at younger ages are subject to speculation. It is clear that HPV infection and low-grade dysplastic lesions are more prevalent at younger ages, but detection and treatment of these lesions appear not to affect the risk of cervical cancer or death to the same degree as at a later age. The most plausible explanation is that the natural history of these lesions differs by age. In this scenario, most cases of CIN2+ detected at a young age would not progress to invasive disease before the next screening invitation. The rare cases that do progress develop rapidly and offer only small possibilities of detection during their short precancerous stage (Sasieni et al. 2009).

Similar patterns of age-specific effect were seen in the mortality data. Participation in screening at the ages of 25–39 reduced the risk of death by 30%, but the estimate was non-significant. In contrast, significant effect estimates of 67% and 71% were observed for participation at 40–54 and 55–69. Some previous studies have reported screening effects on mortality, but none have investigated age-specific effects. Generally, the effect on mortality seems to be greater, probably because of the additional effect of likely diagnosis at an earlier stage in the disease in the screened population. A study from Scotland reported an OR of 4.0 (2.1–8.5) for the association of death from cervical cancer and no previous screening tests as compared with 1–2 previous screening tests and a corresponding OR for the association with cervical cancer diagnosis of 2.8 (2.0–4.1) (Macgregor et al. 1994). A Japanese study with data from Osaka observed an OR of 0.22 (0.03–1.95) for the association between death from cervical cancer and screening participation in the 10 years preceding diagnosis and a corresponding OR of 0.41 (0.13–1.29) for cancer incidence (Sobue et al. 1990). At least one earlier cohort study found a similar relationship between effects on incidence and mortality (Magnus et al. 1987).

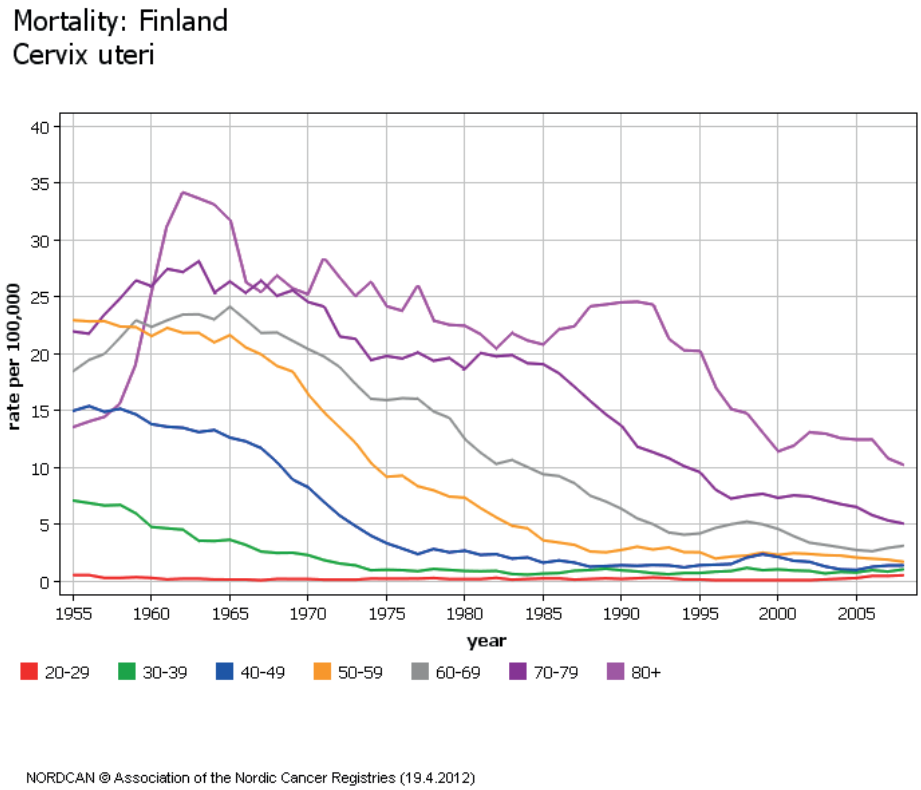

Figure 7.

Reduction in mortality after introduction of the screening programme does not show such clear differences as can be seen in the incidence trends. Large declines can be observed across all ages (see Figure 7). The difference is found in the start of the downward trend, which occurs progressively later as the age band increases. Mortality is affected not only by prevention of cancer by screening but also by better treatment and management, along with earlier diagnosis.

10.5.2 SELF-SELECTION BIAS

The results of the current study point to the large difference that correction for self-selection bias can make in the observed risk reductions associated with participation in screening. The estimates of this bias were based mainly on results for those invited to screening at the age of 25 and 65, and to a lesser degree those invited at 30 and 60. These are the ages where invitation coverage has been variable and both uninvited and invited women were found. Ages 35–55 have been covered well by screening invitations, and the contribution of these age groups to the estimate of the self-selection bias was therefore limited. There were not enough cases in the separate cells for exposure+outcome combinations for age-specific estimation of the selection bias, but if there are large differentials in this bias, the comparison of screening effect between different ages at exposure could be jeopardised. In addition to overall corrections, we produced age-specific self-selection factors by adjusting for age-specific participation rates. This resulted in a smaller correction at young ages when participation has been lower and a larger correction at the ages of 55 and 60 when participation in the Finnish programme has been above average. However, using age-specific self-selection factors did not significantly alter the results: the screening effect was still clearly smaller for women under 40 and was small or absent for screening at 25.

The magnitude of the self-selection bias can be compared to that in early cohort studies for which uninvited populations were available. A Finnish study compared non-participating women's results with expected risks from the time before screening and found a relative risk of 1.6 (Hakama and Rasanen-Virtanen 1976). Also, a Norwegian study observed a relative risk of 1.6 in invited but unscreened women in comparison to neighbouring regions without screening (Magnus et al. 1987). A third cohort study used data from the programme in British Columbia and observed a relative risk of 1.1 for unscreened women (Fidler et al. 1968). Our estimate for the risk of non-participants as compared to non-invited women was 1.29 for stage-IA+ cancers and 1.45 for IB+ cancers. These values are somewhat lower than the ratio found in the cohort study from Finland, but the setting also is very different. Participation rates in the early programme were much higher (85%) than during our

study period. If most of the populace participates in screening, the small proportion not participating may well be a particularly high-risk group of women. If, on the other hand, the group of non-participants is large and heterogeneous, the difference in risk may be smaller. The risk profile of the population was also different from that in the current setting, in which some of the women failing to participate in the index screen may have participated in previous screens, and also overall risks in the population are lower. Opportunistic screening occurs also in the nonparticipating part of the population and may dilute not only the effect of screening but also the baseline risk ratio, because of self-selection into programme participants and non-participants. Indeed, lifetime coverage of the Pap test has been estimated at 98% (Anttila and Nieminen 2007) when opportunistic smears are included.

The self-selection factor for estimated effects on mortality was derived from the incidence data on stage-IB+ cancers. We could not estimate a self-selection factor specifically for mortality, since there were very few deaths with a diagnosis after invitation at ages 25–29 and 65–69. We excluded the microinvasive carcinomas from this analysis, since they contribute very little to cervical cancer mortality.

Case–control studies estimate effectiveness as reduction in the risk of disease associated with the intervention – in this case, participation in screening. Participation in multiple screening events over a lifetime will be associated with incremental reductions in risk, as the duration of lowered risk usually exceeds the screening interval. In earlier studies, exposure has often been defined as ever-screened versus never-screened, and here the total of the incremental effects may be visible. In contrast, many of the more recent studies, ours included, seek to estimate effectiveness as the risk reduction associated with a single participation event in a narrower window of exposure. In the latter case, the estimate will be the average risk reduction effected by that one screening event, in a population with various types of previous screening history. In an established screening programme, the amount of previous screening is necessarily correlated with age: at older ages, those who participate in screening probably have a larger number of previous screens and the potential for any additional incremental gains may be smaller than amongst younger women, all other things being equal. Therefore, it is possible that a valid comparison of screening effects across age groups would require corrections of self-selection bias that are based on directly observed age-specific risk ratios for those not participating as compared to those not invited. In this study, we were not able to use this method, on account of a lack of power, and had to rely on observed overall self-selection factors and age-specifically adjusted factors based on the observed participation rates only. No other studies have attempted to quantify age-specific self-selection factors by the methods we have employed, much less by direct observation, so it is difficult to say whether – and, if so, how much – the age-group comparison would be affected.

10.5.3 DURATION OF EFFECT

The duration of any protective effects of screening participation was explored with definition of exposure as participation in the screening round before index. With a screening interval of five years, the window of exposure thus was located 5–10 years before diagnosis. As expected, the observed reduction in incidence and mortality was less than in the main analysis with exposure defined as participation in the index screen. However, the risk reductions were still clear and significant overall, with a larger reduction in mortality than in incidence. The relationship of effect estimates for incidence and mortality was the same with both definitions of exposure (ratio of ORs: 1.6). Also, the decline in effect over time was similar between the two outcomes, with the value of the OR for participation 5–10 years before diagnosis divided by the OR for participation 0–5 years before diagnosis being 1.4 for both incidence and mortality. However, the duration of the protective effect differed between age groups. In the oldest age group, with invitation at ages 55–69, there was very little to no decline in the reduction in risk over the time period examined, suggesting that screening at this age offers a long-lasting protective effect. Screening at the ages of 40–54 yielded point estimates below unity, but the results were non-significant and the difference with the effect of index screen exposure was clear. In the youngest age group, no remaining effect was observed.

An Italian case–control evaluation also found shorter protection for women under 40 than for women 40 or over when exposure was defined as any smear (positive or negative) and duration as either less than or more than three or five years before diagnosis (Zappa et al. 2004). A case–control study from the UK followed the evolution of odds ratios after an operationally negative smear for up to 6.5 years (Sasieni et al. 2003). There were clear differences in the development of risk after the negative smear in the UK study. Risk reached unity after three years for women aged 20–39, approached the background risk level after five years for women aged 40–54, and remained at a significant 0.5 at six years for women aged 55–69. Their results are very much in line with ours with respect to the duration of effect, even though the approach of using negative smears as exposure does not, in fact, describe screening effect, as discussed in the next section.

10.5.4 DURATION OF LOW RISK AFTER NEGATIVE SCREENING RESULTS

The duration of low risk after a negative screen in the Finnish setting has been assessed previously with screening information from 1971–1976 (Viikki et al. 1999). That cohort study examined the risk of CIN3 and cancer and found standardised incidence ratios of 0.5 to 0.6 with follow-up until 1994. The lowered relative risk

associated with a negative smear was found to become even lower with advancing age, which is consistent with our findings.

Even though studies of screening efficacy and effectiveness have sometimes defined exposure in terms of negative screening tests, that approach is not conceptually unproblematic. The only way in which screening can produce an effect is by leading to the treatment of precancerous lesions and thereby halting the development of cancer. For this to happen, the screening test must detect the abnormality – i.e., be positive. Hence, only a positive screening test can have an impact on cancer incidence and mortality. The risk reduction observed after a negative screening test is caused by the selection of a low-risk population. When reduced risk after a negative screening test is observed, the fact provides evidence of the discriminatory properties of the test, which, in turn, can be assumed to result in the correct identification and treatment of women at higher risk. This approach to evaluating effect is indirect, and many methodology-focused papers have advised against it (Knox 1991, Weiss 1994, Cronin et al. 1998, Zappa and Ciatto 2000). However, the approach is appropriate if the objective is to evaluate the negative predictive value of a screening test – for example, for the estimation of appropriate screening intervals. In this case, selection bias is not an issue, since selection is precisely the parameter of interest.

Our results for the duration of low risk after a negative screen can be compared to those reported from the UK (Sasieni et al. 2003), where the results are presented for age groups similar to ours. The evolution of the OR over time in the various age groups is also very similar between these two studies; perhaps a slightly better NPV is suggested in the Finnish programme, particularly for the 25–39 and 40–54 age groups, but differences are small. These results prompted the authors of the UK study to propose triennial screening for women aged 25–49 and screening every five years for women of 50 and over. Similar policy considerations were applied in the Swedish programme. Even though our results also corroborate a shorter duration of low risk after a negative screen in young women, this does not necessarily mean that the associated age group should be tested more intensively. If little or no protective effect of screening can be demonstrated in this age group, an intensified screening frequency will only increase adverse effects without providing additional protection. For this reason, it is extremely important not to base policy decisions on the NPV of screening alone. One should also consider real effectiveness estimates and balance the potential benefits and harms. Compared with increasing the screening frequency for young women with their high prevalence and turnover of HPV infections and intraepithelial lesions, prophylactic vaccination probably provides a better avenue for addressing cervical cancer risk in women in the younger age groups (Yamamoto et al. 2012).

10.5.5 SCREENING AT THE AGE OF 65

The currently recommended target ages for screening in Finland are 30–60, and these ages are now well covered by invitations at five-year intervals (see Table 5). Some municipalities have also chosen to invite women at the age of 25 and 65, providing us with an opportunity to evaluate the effect of screening participation also at these ages. As the burden of both incidence and mortality lies to a large extent with the ages after the last recommended screening round, screening effectiveness at 65 is of particular interest. A similar effect on incidence in the following five-year period was observed for participation in a single screen at 65 as compared to that of participation at 60, but the small numbers meant that the confidence intervals included 1. In an attempt to include more observations in the estimate, we also analysed the effect of participation at 65 on the risk of cancer or death at any point beyond that age. The resulting corrected OR of the association with cervical cancer was 0.66 (0.29–1.50) for incidence and 0.25 (0.03–2.20) for death, indicative of an additional effect but still non-significant. We further enhanced the observations in the analysis by comparing participants to those not participating, regardless of invitation status. When screening participation at 55, 60, or 65 was compared to no screening participation at 55 or later, significant ORs were observed for both incidence and mortality for screens at 60 and 65. In addition, there was a trend of increasing effect with later screens. We observed risk reductions of 59% (OR: 0.41, CI: 0.23–0.74) and 62% (OR: 0.38, CI: 0.16–0.90) for incidence and mortality associated with a screen at 65 and diagnosis at the ages of 66–80. The effects in comparison to no screen at 55 or later were even larger if the outcome was defined as age at diagnosis of 71–80. This is the age band for which the frequency of cervical cancers currently peaks.

The duration of low risk of cervical intraepithelial lesions and cancer is longer after a negative HPV test than it is after negative cytology (Castle et al. 2012). This period is also age-dependent, such that negative HPV tests predict low risk for a longer period of time at older ages (Schiffman et al. 2011). In addition, HPV prevalence falls with age, while the proportion of cytological changes that are reactive may increase (Leinonen et al. 2009). These facts could provide a reasonable foundation for a so-called exit test after which only those with positive HPV tests would remain in the programme.

10.6 STRENGTHS AND LIMITATIONS

The Finnish setting provides several major advantages for the evaluation of screening activities and impacts. Comprehensive national registers are available for accurate linkage by unique personal identifier between data on screening invitations,

tests, and verifications in the screening register and the outcomes for the whole population in the cancer register. The registers have good coverage and accuracy of information for this purpose. Also, the programme is well-established; the whole population has been uniformly covered.

There was little possibility of verification bias in the cytology audit, as all smears were evaluated by two raters. The review of smears was performed blinded to case status and previous cytodiagnosis, and case smears were randomly mixed with control smears. Controls ensured additionally that information on the specificity of rereading and also on the gold standard was available. It is possible that, despite blinding, the knowledge of reviewing audit smears may have led to increased vigilance and possibly over-calling relative to a normal screening situation. The LSIL+ proportion of controls in the original laboratory results was 1.5%, not very different from the 1.0% seen in routine practice. These values suggest a small loss of specificity (perhaps with a concomitant increase in sensitivity) at most. Correlation of review results between screening and reference laboratories was moderate. It seems that an expert panel's review is needed, to establish a good reference standard for cytology. Single-reference-laboratory review is not reliable, because of the inherent interobserver variability of cytodiagnosis.

It is important to ensure that cases and controls have equal opportunity for screening exposure in the evaluation of screening effectiveness. The Finnish screening programme offers an advantage in this respect, since invitations are sent mainly to women of an age divisible by 5, and all eligible women are invited at these ages, regardless of any preceding screening tests. In many previously published studies, screening exposure has been defined with the exclusion of a 6–12-month period before diagnosis, in order to exclude any diagnostic smears, or smears taken because of symptoms. Because we used only invitational smears as exposure, we did not need to make this exclusion. In addition, we were able to identify screens leading to the detection of cancer and could then index the screening exposure to the preceding five-yearly screening invitation event for the screen-detected case and associated controls.

For the first time in a case–control study for the evaluation of cervical cancer screening effectiveness, corrections for self-selection bias were applied. We found the bias to have a considerable impact on the point estimates. We were also able to adjust the overall selection bias for age-specific participation rates so that comparison of effectiveness between age groups was more reliable. However, we were not able to measure the selection bias directly in the individual age groups, mainly because of the high coverage of the core target ages of the programme. We were also forced to use a selection bias estimated for stage-IB+ cancer as outcome in the corrections for mortality effect, again on account of the small numbers of observations when we used death as outcome. For these reasons, even though it is clear

that participation in screening does cause selection into participants at a lower baseline risk and non-participants at a higher baseline risk, the point estimates for this selection must be considered approximate. However, we are confident that the corrections used are sufficiently accurate not to invalidate our findings.

Opportunistic smears represent the main limitation in this setting. It has been estimated that two thirds of all screening tests are opportunistic. These smears are not yet centrally registered and cannot currently be taken into account in the evaluation of the effectiveness of screening. Hence, any effectiveness estimates are strictly descriptive of organised-screening participation against a backdrop of considerable opportunistic activity. For example, the observation that participation in the organised screening programme at age 25 does not influence the risk of cancer or death does not necessarily mean that the screening test overall has no effect at this age. It is possible that the effect was partly obscured by opportunistic screening of non-participants. However, also women 40 and above use extensive opportunistic screening services and yet large risk reductions were seen after programme participation at these ages. A recent report commissioned by THL on the burden and prevention of HPV disease showed that coverage of invitational-screening non-participants by opportunistic smears decreased with increasing age (THL 2011). On the other hand, the proportion of participants in programme screening who were covered by opportunistic screening tests in a fiveyear period was fairly constant (50%) across the invited age groups. It is clear that differential intensity of opportunistic screening cannot entirely account for the large differences observed in programme screening's effectiveness across age groups.

10.7 SUMMARY AND IMPLICATIONS

Data on screen-detected cervical lesions were compared across three health-care registers: the mass screening register, the cancer register, and the hospital discharge register. There was considerable agreement over the three registers despite their different uses and data collection methods. The mass screening register data on histological diagnoses have high coverage and are useful for statistical and monitoring purposes related to the screening programme. However, for use of the most accurate information available, there are grounds for considering the collection of data on diagnoses and management through systematic data retrieval that involves linkage with other health-care registers. The cancer register data on cervical cancers demonstrated a very high level of completeness, but additional information might be gained through consultation of the other registers with respect to CIN3/AIS lesions.

Only a small proportion of CIN3+ cases, cervical cancers in particular, were due to analytical screen failure. The validity of the cytodiagnosis in the screening

laboratories involved in the screening programme was satisfactory in the setting of very low cancer incidence and mortality having been reached in the target population and low referral rates overall. However, the reproducibility of the cytology is only moderate, and care must be taken so that specificity variations do not produce inequalities in adverse effects and in the cost of screening between regions served by different laboratories. Large variations have been observed in the burden of cancer of the cervix between countries and in historical trends. Hence, regular and comprehensive audits should be performed to monitor the screening process and optimise the performance of cytology.

Large variations by laboratory were observed in performance indicators describing the screening process. Despite the apparent differences in reporting policies, no significant variation in the rates of missed progressive lesions could be demonstrated. The programme functions with a good and comparable effect across geographical regions. This means that the monitoring of performance indicators alone is not enough; evaluations and probably also audits of longitudinal outcome measures are needed for ensuring a screening service of high quality. Information provided by performance indicators is important too, and feedback on reporting criteria for cytology is clearly indicated if we are to harmonise the screening service and avoid adverse effects and unnecessary costs.

A single screening episode within the organised screening programme reduces the risk of cervical cancer by half. The reduction in the risk of death from cervical cancer is even larger. However, this effect is age-dependent such that screening participation at ages below 40 is associated with a clearly smaller risk reduction than participation at age 40 and above. No incidence effect was observed at the age of 25, but effect on mortality could not be estimated, because of the small numbers involved. Opportunistic screening is, and has been, in widespread use in the population, especially at younger ages, and, in order for final conclusions to be drawn on the effect of screening at ages below 30, evaluations that take these tests too into consideration are needed.

The duration of reduced risk was also age-dependent and increased with age. It seems that screening participation at age 65 would yield additional incidence and mortality benefits at the ages at which advanced disease and death from cervical cancer are currently most frequent.

The remaining cervical cancer cases and deaths in a well-screened population occur to a large extent more than five years after the last screening invitation. Cancers and, especially, deaths from cervical cancer with a diagnosis before first invitation, are rare. These observations in combination with the findings on age-specific effectiveness may be useful for the planning of future modifications to the screening programme. Another group of concern is that of the invited nonattenders, which contributes a quarter to a third of all cancer cases and deaths. Continuing

efforts are needed to ensure that the screening programme reaches in particular that currently noncompliant group of women that has been shown to demonstrate a higher baseline risk of cervical cancer. Cancers arising after negative screening tests constitute a smaller proportion of cancers and cancer deaths, but some benefit could also be gained through improved sensitivity of the screening test, provided that specificity can be maintained.

11 CONCLUSIONS

- The quality and especially the coverage of the screening register with respect to histopathological diagnoses of screen-detected lesions are sufficiently high for the production of statistics and monitoring data. However, the accuracy of diagnostic information can be improved by data retrieval from other health-care registers. In particular, linkage to the cancer register is important for accurate individual data on cervical cancers.

- The effect of analytical screening failures on cancer incidence is fairly small and cannot easily be further decreased without reduction in the high overall specificity of screening.

- Considerable variations exist in the balance between the sensitivity and specificity of screening laboratories, but these variations are not reflected in screening effectiveness. Regular feedback is needed, to improve the reproducibility of the cytodiagnosis in the long term and harmonise the screening service with respect to cost and the physical and psychological burdens caused by the management of abnormal screening tests.

- The effectiveness of organised screening is strongly age-dependent such that programme participation at the age of 25 is not associated with a reduced risk of cervical cancer in the following screening interval. Effectiveness seems to persist until the age of 65.

- A large proportion of cervical cancer incidence and most of the mortality are due to cancers diagnosed more than five years after the last screening invitation, the next largest group consists of non-attenders, and a smaller proportion of cancers are diagnosed among women who have attended invitational screening.

Even though there are sensitivity issues in cytology, they are alleviated by the combination of regularly repeated tests and the long screen-detectable phase in the development of cervical cancer. Our study has shown within the Finnish programme setting that most women who develop cervical cancer do so because of lack of screening rather than on account of errors in the cytodiagnosis. This observation should direct further efforts to improvement in the quality assurance and effectiveness of cervical cancer screening programmes.

12 ACKNOWLEDGEMENTS

I would like to thank my principal guides along the way, Docent Ahti Anttila and Docent Pekka Nieminen. Together they are responsible for my introduction and gradual integration into the field of cervical cancer screening. I am very pleased to have had the opportunity to work with Ahti Anttila. He has been more than generous with his time, trust, and broad experience and understanding of the complexities of screening. I am also very pleased to have had the support of the heavy clinical experience of Pekka Nieminen, and almost equally importantly, his irrepressible and contagious optimism and drive.

I want to thank Professor Jaakko Kaprio, at the Department of Public Health, for support during the stages leading up to completion, publication and disputation.

I would like to thank Professor Timo Hakulinen, Director of the Finnish Cancer Registry, a research institute I have very much enjoyed working at these past few years. Professor Nea Malila has been, besides a co-author, my immediate administrative superior and has ensured an excellent and appreciative working environment at the Mass Screening Registry. That working environment has also been greatly enhanced by Anni Pehkonen, Tiina Karhunen, Kaija Halonen, Minna Heikkilä, Sanni Helander, Liisa Määttänen, Tero Näveri, Sanna Kuivalainen, Liisa Rita, Päivi Styrman, Laura Madanat-Harjuoja, Tytti Sarkeala, and Jonna Fredman. Professor Emeritus Matti Hakama has provided scientific guidance and maybe even more importantly, inspiration.

Maarit Leinonen, Ilkka Kalliala and Anni Virtanen have been part of the cervical cancer scientific team, working with different aspects of the same whole; I appreciate the peer support you have provided.

Statistician Tapio Luostarinen has been an important scientific collaborator throughout. I have also enjoyed collaborating with the other co-authors: Laura Kotaniemi-Talonen, Harry Kujari, Jukka Melkko, Gustav Granroth, Martine Vornanen, Timo Pietiläinen, Anna Sankila, and Johanna Arola.

I would also like to thank the expert reviewers of the thesis, Professor Johanna Mäenpää and Docent Simopekka Vänskä for their knowledgeable and constructive criticism. Anu Planting commented on the language of this thesis and her fast and thorough work is greatly appreciated.

There are some that have not directly participated in the work but still unwittingly provided inspiration to the work at hand. These people include at least Lawrence von Karsa, Sven Törnberg and Guglielmo Ronco.

Thank you Johanna for your love, support, flexibility, wit and wisdom, all of which are invaluable attributes in a life partner, maybe especially during projects such as this. Erik and Ylva, this is to you.

The financial support of the Finnish Cancer Organisations, the Doctoral School of Public Health, the Finnish Medical Society Duodecim and the European Union through the action programme Europe Against Cancer and the EUROCOURSE project is gratefully acknowledged.

Oslo, September 2012

Stefan Lönnberg

13 REFERENCES

Aklimunnessa K, Mori M, Khan MM, Sakauchi F, Kubo T, Fujino Y, Suzuki S, Tokudome S, Tamakoshi A, Motohashi Y, Tsuji I, Nakamura Y, Iso H, Mikami H, Inaba Y, Hoshiyama Y, Suzuki H, Shimizu H, Toyoshima H, Wakai K, Ito Y, Hashimoto S, Kikuchi S, Koizumi A, Kawamura T, Watanabe Y, Miki T, Date C, Sakata K, Nose T, Hayakawa N, Yoshimura T, Shibata A, Okamoto N, Shino H, Ohno Y, Kitagawa T, Kuroki T, and Tajima K (2006) Effectiveness of cervical cancer screening over cervical cancer mortality among Japanese women. Jpn J Clin Oncol 36**:** 511-8.

Anderson GH, Benedet JL, Le Riche JC, Matisic JP, and Thompson JE (1992) Invasive cancer of the cervix in British Columbia: a review of the demography and screening histories of 437 cases seen from 1985-1988. Obstet Gynecol 80**:** 1-4.

Andersson E, Karrberg C, Radberg T, Blomqvist L, Zetterqvist BM, Ryd W, Lindh M, and Horal P (2012) Type-dependent E6/E7 mRNA expression of single and multiple high-risk human papillomavirus infections in cervical neoplasia. J Clin Virol 54**:** 61-5.

Andersson-Ellström A, Seidal T, Grannas M, and Hagmar B (2000) The Pap-smear history of women with invasive cervical squamous carcinoma. Acta Obstet Gynecol Scand 79**:** 221-6.

Andrae B and Smith P (1999) Clinical impact of quality assurance in an organized cervical screening program. Acta Obstet Gynecol Scand 78**:** 429-35.

Andrae B, Strander B, Silfverdal L, Ryd W, Dillner J, Törnberg S, and Sparen P (2009) Benefit of cervical cancer screening in young women – a matter of adherence to the recommended screening interval [Rapid Response]. BMJ. http://www.bmj.com/

Andrae B, Kemetli L, Sparen P, Silfverdal L, Strander B, Ryd W, Dillner J, and Törnberg S (2008) Screening-preventable cervical cancer risks: evidence from a nationwide audit in Sweden. J Natl Cancer Inst 100**:** 622-9.

Andrae B, Andersson TM, Lambert PC, Kemetli L, Silfverdal L, Strander B, Ryd W, Dillner J, Törnberg S, and Sparén P (2012) Screening and cervical cancer cure: population based cohort study. BMJ 344**:** e900.

Anttila A and Nieminen P (2000) Cervical cancer screening programme in Finland. Eur J Cancer 36**:** 2209-14.

Anttila A and Nieminen P (2007) Cervical cancer screening programme in Finland with an example on implementing alternative screening methods. Coll Antropol 31 Suppl 2**:** 17-22.

Anttila A, Hakama M, Kotaniemi-Talonen L, and Nieminen P (2006) Alternative technologies in cervical cancer screening: a randomised evaluation trial. BMC Public Health 6**:** 252.

Anttila A, Kotaniemi-Talonen L, Leinonen M, Hakama M, Laurila P, Tarkkanen J, Malila N, and Nieminen P (2010) Rate of cervical cancer, severe intraepithelial neoplasia, and adenocarcinoma in situ in primary HPV DNA screening with cytology triage: randomised study within organised screening programme. BMJ 340**:** c1804.

Anttila A, von Karsa L, Aasmaa A, Fender M, Patnick J, Rebolj M, Nicula F, Vass L, Valerianova Z, Voti L, Sauvaget C, and Ronco G (2009) Cervical cancer screening policies and coverage in Europe. Eur J Cancer 45**:** 2649-58.

Arbyn M, Raifu AO, Weiderpass E, Bray F, and Anttila A (2009b) Trends of cervical cancer mortality in the member states of the European Union. Eur J Cancer 45**:** 2640-8.

Arbyn M, Bergeron C, Klinkhamer P, Martin-Hirsch P, Siebers AG, and Bulten J (2008b) Liquid compared with conventional cervical cytology: a systematic review and meta-analysis. Obstet Gynecol 111**:** 167-77.

Arbyn M, Rebolj M, De Kok IM, Fender M, Becker N, O'Reilly M, and Andrae B (2009a) The challenges of organising cervical screening programmes in the 15 old member states of the European Union. Eur J Cancer 45**:** 2671-8.

Arbyn M, Kyrgiou M, Simoens C, Raifu AO, Koliopoulos G, Martin-Hirsch P, Prendiville W, and Paraskevaidis E (2008c) Perinatal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: meta-analysis. BMJ 337**:** a1284.

Arbyn M, Anttila A, Jordan J, Ronco G, Schenck U, Segnan N, Wiener H, Herbert A, Daniel J, and von Karsa L (eds) (2008a) European guidelines for quality assurance in cervical cancer screening. Office for Official Publications of the European Communities: Luxembourg.

Austin RM and Zhao C (2012) Type 1 and type 2 cervical carcinomas: some cervical cancers are more difficult to prevent with screening. Cytopathology 23: 6-12.

Auvinen E, Niemi M, Malm C, Zilliacus R, Trontti A, Fingerroos R, Lehtinen M, and Paavonen B (2005) High prevalence of HPV among female students in Finland. Scand J Infect Dis 37: 873-6.

Bagnall H, Pearmain P, Clare J, and Lawrence G (2006) A new method for the classification of invasive cervical cancer screening histories. J Med Screen 13: 137-47.

Barrios K and Celis E (2012) TriVax-HPV: an improved peptide-based therapeutic vaccination strategy against human papillomavirus-induced cancers. Cancer Immunol Immunother 61: 1307-17.

Berkhof J, de Bruijne MC, Zielinski GD, and Meijer CJ (2005) Natural history and screening model for high-risk human papillomavirus infection, neoplasia and cervical cancer in the Netherlands. Int J Cancer 115: 268-75.

Blanks RG (2008) The development of the referral outcome diagram and an analysis of laboratory cancer detection rates in the English NHS cervical screening programme – is there an optimum level of detection of CIN 1 and CIN 2 lesions? Cytopathology 19: 244-53.

Bos AB, Rebolj M, Habbema JD, and van Ballegooijen M (2006) Nonattendance is still the main limitation for the effectiveness of screening for cervical cancer in the Netherlands. Int J Cancer 119: 2372-5.

Bosch FX, Manos MM, Muñoz N, Sherman M, Jansen AM, Peto J, Schiffman MH, Moreno V, Kurman R, and Shah KV (1995) Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. J Natl Cancer Inst 87: 796-802.

Brotherton JM, Fridman M, May CL, Chappell G, Saville AM, and Gertig DM (2011) Early effect of the HPV vaccination programme on cervical abnormalities in Victoria, Australia: an ecological study. Lancet 377**:** 2085-92.

Brown CA, Bogers J, Sahebali S, Depuydt CE, De Prins F, and Malinowski DP (2012) Role of protein biomarkers in the detection of high-grade disease in cervical cancer screening programs. J Oncol 2012**:** 289315.

Brown RK and Barker WH, Jr. (1982) Pap smear screening and invasive cervical cancer. J Fam Pract 15**:** 875-9.

Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, and Meijer GA (2007) Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 29**:** 19-24.

Castañon A, Ferryman S, Patnick J, and Sasieni P (2012) Review of cytology and histopathology as part of the NHS Cervical Screening Programme audit of invasive cervical cancers. Cytopathology 23**:** 13-22.

Castle PE, Glass AG, Rush BB, Scott DR, Wentzensen N, Gage JC, Buckland J, Rydzak G, Lorincz AT, and Wacholder S (2012) Clinical human papillomavirus detection forecasts cervical cancer risk in women over 18 years of follow-up. J Clin Oncol 30**:** 3044-50.

Chamberlain J (1984) Failures of the cervical cytology screening programme. Br Med J (Clin Res Ed) 289**:** 853-4.

Coleman DV and Poznansky JJ (2006) Review of cervical smears from 76 women with invasive cervical cancer: cytological findings and medicolegal implications. Cytopathology 17**:** 127-36.

Crocetti E, Battisti L, Betta A, Palma PD, Paci E, Piffer S, Pojer A, Polla E, and Zappa M (2007) The cytological screening turned out effective also for adenocarcinoma: a population-based casecontrol study in Trento, Italy. Eur J Cancer Prev 16**:** 564-7.

Cronin KA, Weed DL, Connor RJ, and Prorok PC (1998) Case-control studies of cancer screening: theory and practice. J Natl Cancer Inst 90**:** 498-504.

Cuzick J (2008) Routine audit of large-scale cervical cancer screening programs. J Natl Cancer Inst 100: 605-6.

Cuzick J, Arbyn M, Sankaranarayanan R, Tsu V, Ronco G, Mayrand MH, Dillner J, and Meijer CJ (2008) Overview of human papillomavirus-based and other novel options for cervical cancer screening in developed and developing countries. Vaccine 26 Suppl 10: K29-41.

de Bie RP, Vergers-Spooren HC, Massuger LF, Siebers AG, Salet-van der Pol MR, Vedder JE, Melchers WJ, Bulten J, and Bekkers RL (2011) Patients with cervical cancer: why did screening not prevent these cases? Am J Obstet Gynecol 205: 64 e1-7.

de Kok IM, van Rosmalen J, Dillner J, Arbyn M, Sasieni P, Iftner T, and van Ballegooijen M (2012) Primary screening for human papillomavirus compared with cytology screening for cervical cancer in European settings: cost effectiveness analysis based on a Dutch microsimulation model. BMJ 344: e670.

de Villiers EM, Fauquet C, Broker TR, Bernard HU, and zur Hausen H (2004) Classification of papillomaviruses. Virology 324: 17-27.

Duffy SW, Cuzick J, Tabar L, Vitak B, Chen TH-H, Yen M-F, and Smith RA (2002) Correcting for Non-compliance bias in case-control studies to evaluate cancer screening programmes. J Roy Statl Soc C-App 51: 235-243.

Dunn JE, Jr., and Schweitzer V (1981) The relationship of cervical cytology to the incidence of invasive cervical cancer and mortality in Alameda County, California, 1960 to 1974. Am J Obstet Gynecol 139: 868-76.

EC (2003) Council recommendation of 2 December 2003 on cancer screening (2003/878/EC) OJ L of 16.12.2003: 34-8.

Einstein MH, Baron M, Levin MJ, Chatterjee A, Fox B, Scholar S, Rosen J, Chakhtoura N, Meric D, Dessy FJ, Datta SK, Descamps D, and Dubin G (2011) Comparative immunogenicity and safety of human papillomavirus (HPV)-16/18 vaccine and HPV-6/11/16/18 vaccine: follow-up from months 12-24 in a Phase III randomized study of healthy women aged 18-45 years. Hum Vaccin 7: 1343-58.

Ejersbo D (2008) [Screening profile of women who have died from cervical cancer]. Ugeskr Laeger 170: 727-30.

ENCR (2001) Method of Detection in Relation to Screening. www.encr.com.fr.

Engholm G, Ferlay J, Christensen N, Johannesen TB, Klint Å, Køtlum JE, Milter MC, Ólafsdóttir E, Pukkala E, Storm HH. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 5.1 (March 2012). Association of the Nordic Cancer Registries. Danish Cancer Society. http://www.ancr.nu.

Fidler HK, Boyes DA, and Worth AJ (1968) Cervical cancer detection in British Columbia: a progress report. J Obstet Gynaecol Br Commonw 75: 392-404.

Finnish Cancer Registry (2009) Cancer in Finland 2006 and 2007. Cancer Society of Finland Publication No. 76. Cancer Society of Finland: Helsinki.

Finnish Cancer Registry (2011) Cancer in Finland 2008 and 2009. Cancer Society of Finland Publication No. 84. Cancer Society of Finland: Helsinki.

Gravitt PE, Coutlee F, Iftner T, Sellors JW, Quint WG, and Wheeler CM (2008) New technologies in cervical cancer screening. Vaccine 26 Suppl 10: K42-52.

Grundsell H, Johnsson JE, Lindberg LG, Strom H, Tekavec E, Trope C, and Bekassy Z (1979) Vaginal smear history in patients with invasive cervical carcinoma. Ann Chir Gynaecol 68: 127-9.

Gwet K (2002) Inter-Rater Reliability: Dependency on Trait Prevalence and Marginal Homogeneity. Statistical Mehtods For Inter-Rater Reliability Assessment 2: 1-9.

Hakama M and Räsänen-Virtanen U (1976) Effect of a mass screening program on the risk of cervical cancer. Am J Epidemiol 103: 512-7.

Hakama M, Chamberlain J, Day NE, Miller AB, and Prorok PC (1985) Evaluation of screening programmes for gynaecological cancer. Br J Cancer 52: 669-73.

Hellsten C, Sjostrom K, and Lindqvist PG (2008) A 2-year follow-up study of anxiety and depression in women referred for colposcopy after an abnormal cervical smear. BJOG 115: 212-8.

Herbert A, Anshu, Gregory M, Gupta SS, and Singh N (2009a) Invasive cervical cancer audit: a relative increase in interval cancers while coverage increased and incidence declined. BJOG 116**:** 845-53.

Herbert A, Anshu, Gregory M, Gupta SS, and Singh N (2009b) Screen-detected invasive cervical carcinoma and its clinical significance during the introduction of organized screening. BJOG 116**:** 854-9.

Herbert A, Anshu, Culora G, Dunsmore H, Gupta SS, Holdsworth G, Kubba AA, McLean E, Sim J, and Raju KS (2010) Invasive cervical cancer audit: why cancers developed in a high-risk population with an organised screening programme. BJOG 117**:** 736-45.

Hernandéz-Avila M, Lazcano-Ponce EC, de Rúiz PA, and Romieu I (1998) Evaluation of the cervical cancer screening programme in Mexico: a population-based case-control study. Int J Epidemiol 27**:** 370-6.

Herrero R, Brinton LA, Reeves WC, Brenes MM, de Britton RC, Gaitan E, and Tenorio F (1992) Screening for cervical cancer in Latin America: a case-control study. Int J Epidemiol 21**:** 1050-6.

Hoffman M, Cooper D, Carrara H, Rosenberg L, Kelly J, Stander I, Williamson AL, Denny L, du Toit G, and Shapiro S (2003) Limited Pap screening associated with reduced risk of cervical cancer in South Africa. Int J Epidemiol 32**:** 573-7.

IARC (1986) Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. IARC Working Group on Evaluation of Cervical Cancer Screening Programmes. Br Med J (Clin Res Ed) 293**:** 659-64.

IARC (2005) IARC Handbooks of Cancer Prevention, volume 10: Cervix Cancer Screening. IARC Press: Lyon.

IARC (2007) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, volume 96: Human Papillomaviruses. IARC Press: Lyon.

Ingemann-Hansen O, Lidang M, Niemann I, Dinesen J, Baandrup U, Svanholm H, and Petersen L (2008) Screening history of women with cervical cancer: a 6-year study in Aarhus, Denmark. Br J Cancer 98**:** 1292-4.

Janerich DT, Hadjimichael O, Schwartz PE, Lowell DM, Meigs JW, Merino MJ, Flannery JT, and Polednak AP (1995) The screening histories of women with invasive cervical cancer, Connecticut. Am J Public Health 85**:** 791-4.

Kasinpila C, Promthet S, Vatanasapt P, Sasieni P, and Parkin DM (2011) Evaluation of the nationwide cervical screening programme in Thailand: a case-control study. J Med Screen 18**:** 147-53.

Kenter GG, Schoonderwald EM, Koelma IA, Arentz N, Hermans J, and Fleuren GJ (1996) The cytological screening history of 469 patients with squamous cell carcinoma of the cervix uteri; does interval carcinoma exist? Acta Obstet Gynecol Scand 75**:** 400-3.

Kirschner B, Poll S, Rygaard C, Wahlin A, and Junge J (2011) Screening history in women with cervical cancer in a Danish population-based screening program. Gynecol Oncol 120**:** 68-72.

Kitchener HC, Gilham C, Sargent A, Bailey A, Albrow R, Roberts C, Desai M, Mather J, Turner A, Moss S, and Peto J (2011) A comparison of HPV DNA testing and liquid based cytology over three rounds of primary cervical screening: Extended follow up in the ARTISTIC trial. Eur J Cancer 47**:** 864-71.

Knox G (1991) Case-control studies of screening procedures. Public Health 105**:** 55-61.

Kotaniemi-Talonen L, Nieminen P, Hakama M, Seppänen J, Ikkala J, Martikainen J, Tarkkanen J, Toivonen T, and Anttila A (2007) Significant variation in performance does not reflect the effectiveness of the cervical cancer screening programme in Finland. Eur J Cancer 43**:** 169-74.

Kristensen GB, Skyggebjerg KD, Holund B, Holm K, and Hansen MK (1991) Analysis of cervical smears obtained within three years of the diagnosis of invasive cervical cancer. Acta Cytol 35**:** 47-50.

Laurenson S, Pett MR, Hoppe-Seyler K, Denk C, Hoppe-Seyler F, Coleman N, and Ko Ferrigno P (2011) Development of peptide aptamer microarrays for detection of HPV16 oncoproteins in cell extracts. Anal Biochem 410**:** 161-70.

Lehtinen M, Kaasila M, Pasanen K, Patama T, Palmroth J, Laukkanen P, Pukkala E, and Koskela P (2006) Seroprevalence atlas of infections with oncogenic and non-oncogenic human papillomaviruses in Finland in the 1980s and 1990s. Int J Cancer 119**:** 2612-9.

Lehtinen M, Paavonen J, Wheeler CM, Jaisamrarn U, Garland SM, Castellsagué X, Skinner SR, Apter D, Naud P, Salmerón J, Chow SN, Kitchener H, Teixeira JC, Hedrick J, Limson G, Szarewski A, Romanowski B, Aoki FY, Schwarz TF, Poppe WA, De Carvalho NS, Germar MJ, Peters K, Mindel A, De Sutter P, Bosch FX, David MP, Descamps D, Struyf F, Dubin G; HPV PATRICIA Study Group (2012) Overall efficacy of HPV-16/18 AS04-adjuvanted vaccine against grade 3 or greater cervical intraepithelial neoplasia: 4-year end-of-study analysis of the randomised, double-blind PATRICIA trial. Lancet Oncol 13**:** 89-99.

Leinonen M, Nieminen P, Kotaniemi-Talonen L, Malila N, Tarkkanen J, Laurila P, and Anttila A (2009) Age-specific evaluation of primary human papillomavirus screening vs conventional cytology in a randomized setting. J Natl Cancer Inst 101**:** 1612-23.

Lindqvist PG, Hellsten C, and Rippe A (2008) Screening history of women in Malmö with invasive cervical cancer. Eur J Obstet Gynecol Reprod Biol 137**:** 77-83.

Macgregor JE, Campbell MK, Mann EM, and Swanson KY (1994) Screening for cervical intraepithelial neoplasia in north east Scotland shows fall in incidence and mortality from invasive cancer with concomitant rise in preinvasive disease. BMJ 308**:** 1407-11.

Magnus K, Langmark F, and Andersen A (1987) Mass screening for cervical cancer in Østfold county of Norway 1959-77. Int J Cancer 39**:** 311-6.

Makino H, Sato S, Yajima A, Komatsu S, and Fukao A (1995) Evaluation of the effectiveness of cervical cancer screening: a case-control study in Miyagi, Japan. Tohoku J Exp Med 175**:** 171-8.

McCredie MR, Sharples KJ, Paul C, Baranyai J, Medley G, Jones RW, and Skegg DC (2008) Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. Lancet Oncol 9**:** 425-34.

McCrory DC, Matchar DB, Bastian L, Datta S, Hasselblad V, Hickey J, Myers E, and Nanda K (1999) Evaluation of cervical cytology. Evid Rep Technol Assess (Summ): 1-6.

Miller AB (1995) Failures of cervical cancer screening. Am J Public Health 85: 761-2.

Mitchell H, Hocking J, and Saville M (2003) Improvement in protection against adenocarcinoma of the cervix resulting from participation in cervical screening. Cancer 99: 336-41.

Mitchell H, Medley G, and Giles G (1990) Cervical cancers diagnosed after negative results on cervical cytology: perspective in the 1980s. BMJ 300: 1622-6.

Mitchell H, Medley G, and Higgins V (1996) An audit of the women who died during 1994 from cancer of the cervix in Victoria, Australia. Aust N Z J Obstet Gynaecol 36: 73-6.

Mitchell H, Medley G, Gordon I, and Giles G (1995) Cervical cytology reported as negative and risk of adenocarcinoma of the cervix: no strong evidence of benefit. Br J Cancer 71: 894-7.

Morrison AS (1982) Case definition in case-control studies of the efficacy of screening. Am J Epidemiol 115: 6-8.

Moss SM (1991) Case-control studies of screening. Int J Epidemiol 20: 1-6.

Muñoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders PJ, and Meijer CJ (2003) Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 348: 518-27.

Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, and Matchar DB (2000) Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. Ann Intern Med 132: 810-9.

NHS Cervical Screening Programme (2006) Audit of cervical cancers. NHSCSP Publication No. 28, 1st edn. NHS Cancer Screening Programmes: Sheffield.

Nieminen P, Kallio M, Anttila A, and Hakama M (1999) Organised vs. spontaneous Pap-smear screening for cervical cancer: A case-control study. Int J Cancer 83: 55-8.

Nieminen P, Hakama M, Viikki M, Tarkkanen J, and Anttila A (2003) Prospective and randomised public-health trial on neural network-assisted screening for cervical cancer in Finland: results of the first year. Int J Cancer 103: 422-6.

Nieminen P, Anttila A, Bützow R, Heikkila E, Hiltunen-Back E, Mäenpää J, Puistola U, Rantanen V, Rintala M, Räisanen I, Santalahti A, Talvensaari-Mattila A, Vartiainen J, Vuento M, and Yliskoski M (2010) [Update on Current Care Guidelines: Diagnosis, treatment and follow-up of cytological changes in the cervix, vagina and vulva]. Duodecim 126: 1965-6.

Otto SJ, Fracheboud J, Verbeek AL, Boer R, Reijerink-Verheij JC, Otten JD, Broeders MJ, and de Koning HJ (2012) Mammography screening and risk of breast cancer death: a population-based case-control study. Cancer Epidemiol Biomarkers Prev 21: 66-73.

Pajunen P, Koukkunen H, Ketonen M, Jerkkola T, Immonen-Räiha P, Kärjä-Koskenkari P, Mähönen M, Niemelä M, Kuulasmaa K, Palomäki P, Mustonen J, Lehtonen A, Arstila M, Vuorenmaa T, Lehto S, Miettinen H, Torppa J, Tuomilehto J, Kesäniemi YA, Pyörälä K, and Salomaa V (2005) The validity of the Finnish Hospital Discharge Register and Causes of Death Register data on coronary heart disease. Eur J Cardiovasc Prev Rehabil 12: 132-7.

Palli D, Carli S, Cecchini S, Venturini A, Piazzesi G, and Buiatti E (1990) A centralised cytology screening programme for cervical cancer in Florence. J Epidemiol Community Health 44: 47-51.

Papanicolaou GN and Traut HF (1941) The diagnostic value of vaginal smears in carcinoma of the uterus. Am J Obstet Gynecol 42: 193-206.

Pecorelli S, Zigliani L, and Odicino F (2009) Revised FIGO staging for carcinoma of the cervix. Int J Gynaecol Obstet 105: 107-8.

Percy C, Fritz A, Jack A, Shanmugarathan S, Sobin L, Parkin DM, and Whelan S (eds) (2000) International Classification of Diseases for Oncology (ICD-O). World Health Organization: Geneva.

Renshaw AA, Young ML, and Holladay EB (2004) Blinded review of Papanicolaou smears in the context of litigation. Cancer 102**:** 136-41.

Repše-Fokter A, Pogačnik A, Snoj V, Primic-Žakelj M, and Fležar MS (2012) Review of negative and low-grade cervical smears in women with invasive cervical cancer after the first 3 years of the national cervical screening programme in Slovenia. Cytopathology 23**:** 23-9.

Robertson JH and Woodend B (1993) Negative cytology preceding cervical cancer: causes and prevention. J Clin Pathol 46**:** 700-2.

Rodriguez AC, Burk R, Herrero R, Hildesheim A, Bratti C, Sherman ME, Solomon D, Guillen D, Alfaro M, Viscidi R, Morales J, Hutchinson M, Wacholder S, and Schiffman M (2007) The natural history of human papillomavirus infection and cervical intraepithelial neoplasia among young women in the Guanacaste cohort shortly after initiation of sexual life. Sex Transm Dis 34**:** 494-502.

Ronco G, Giubilato P, Naldoni C, Zorzi M, Anghinoni E, Scalisi A, Dalla Palma P, Zanier L, Barca A, Angeloni C, Gaimo MD, Maglietta R, Mancini E, Pizzuti R, Iossa A, Segnan N, and Zappa M (2010b) Extension of organised cervical cancer screening programmes in Italy and their process indicators: 2008 activity. Epidemiol Prev 34**:** 35-51.

Ronco G, van Ballegooijen M, Becker N, Chil A, Fender M, Giubilato P, Kurtinaitis J, Lancucki L, Lynge E, Morais A, O'Reilly M, Sparen P, Suteu O, Rebolj M, Veerus P, Zakelj MP, and Anttila A (2009) Process performance of cervical screening programmes in Europe. Eur J Cancer 45**:** 2659-70.

Ronco G, Giorgi-Rossi P, Carozzi F, Confortini M, Dalla Palma P, Del Mistro A, Ghiringhello B, Girlando S, Gillio-Tos A, De Marco L, Naldoni C, Pierotti P, Rizzolo R, Schincaglia P, Zorzi M, Zappa M, Segnan N, and Cuzick J (2010a) Efficacy of human papillomavirus testing for the detection of invasive cervical cancers and cervical intraepithelial neoplasia: a randomised controlled trial. Lancet Oncol 11**:** 249-57.

Rylander E (1976) Cervical cancer in women belonging to a cytologically screened population. Acta Obstet Gynecol Scand 55**:** 361-6.

Sankaranarayanan R, Esmy PO, Rajkumar R, Muwonge R, Swaminathan R, Shanthakumari S, Fayette JM, and Cherian J (2007) Effect of visual screening on cervical cancer incidence and mortality in Tamil Nadu, India: a cluster-randomised trial. Lancet 370**:** 398-406.

Sankaranarayanan R, Nene BM, Shastri SS, Jayant K, Muwonge R, Budukh AM, Hingmire S, Malvi SG, Thorat R, Kothari A, Chinoy R, Kelkar R, Kane S, Desai S, Keskar VR, Rajeshwarkar R, Panse N, and Dinshaw KA (2009) HPV screening for cervical cancer in rural India. N Engl J Med 360**:** 1385-94.

Sasieni P and Cuzick J (2001) Routine audit is an ethical requirement of screening. BMJ 322**:** 1179.

Sasieni P, Adams J, and Cuzick J (2003) Benefit of cervical screening at different ages: evidence from the UK audit of screening histories. Br J Cancer 89**:** 88-93.

Sasieni P, Castañon A, and Cuzick J (2009a) Screening and adenocarcinoma of the cervix. Int J Cancer 125**:** 525-9.

Sasieni P, Castañon A, and Cuzick J (2009b) Effectiveness of cervical screening with age: population based case-control study of prospectively recorded data. BMJ 339**:** b2968.

Sasieni PD, Cuzick J, and Lynch-Farmery E (1996) Estimating the efficacy of screening by auditing smear histories of women with and without cervical cancer. The National Co-ordinating Network for Cervical Screening Working Group. Br J Cancer 73**:** 1001-5.

Schiffman M, Wentzensen N, Wacholder S, Kinney W, Gage JC, and Castle PE (2010) Human papillomavirus testing in the prevention of cervical cancer. J Natl Cancer Inst 103**:** 368-83.

Schiffman M, Glass AG, Wentzensen N, Rush BB, Castle PE, Scott DR, Buckland J, Sherman ME, Rydzak G, Kirk P, Lorincz AT, Wacholder S, and Burk RD (2011) Cancer Epidemiol Biomarkers Prev 20**:** 1398-409.

Slater DN, Milner PC, and Radley H (1994) Audit of deaths from cervical cancer: proposal for an essential component of the National Screening Programme. J Clin Pathol 47**:** 27-8.

Sobue T, Suzuki T, Fujimoto I, Yokoi N, and Naruke T (1990) Population-based case-control study on cancer screening. Environ Health Perspect 87**:** 57-62.

Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, Raab S, Sherman M, Wilbur D, Wright T, Jr., and Young N (2002) The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA 287**:** 2114-9.

Spayne J, Ackerman I, Milosevic M, Seidenfeld A, Covens A, and Paszat L (2008) Invasive cervical cancer: a failure of screening. Eur J Public Health 18**:** 162-5.

Spence AR, Goggin P, and Franco EL (2007) Process of care failures in invasive cervical cancer: systematic review and meta-analysis. Prev Med 45**:** 93-106.

Stoler MH and Schiffman M (2001) Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. JAMA 285**:** 1500-5.

Stuart GC, McGregor SE, Duggan MA, and Nation JG (1997) Review of the screening history of Alberta women with invasive cervical cancer. CMAJ 157**:** 513-9.

Sung HY, Kearney KA, Miller M, Kinney W, Sawaya GF, and Hiatt RA (2000) Papanicolaou smear history and diagnosis of invasive cervical carcinoma among members of a large prepaid health plan. Cancer 88**:** 2283-9.

Syrjänen KJ (2009) Histology, classification, and natural history of cervical intraepithelial neoplasia (CIN). CME J Gynecol Oncol 14**:** 4-21.

Teppo L, Pukkala E, and Lehtonen M (1994) Data quality and quality control of a population-based cancer registry. Experience in Finland. Acta Oncol 33**:** 365-9.

THL (2011) Terveyden ja hyvinvoinnin laitoksen asettaman papilloomavirustautien torjuntatyöryhmän selvitys [Report of the study of the Expert Group on Papillomavirus Disease Prevention appointed by the National Institute for Health and Welfare]. THL Report 28/2011.

Tolonen H, Salomaa V, Torppa J, Sivenius J, Immonen-Räihä P, and Lehtonen A (2007) The validation of the Finnish Hospital Discharge Register and Causes of Death Register data on stroke diagnoses. Eur J Cardiovasc Prev Rehabil 14**:** 380-5.

Tsoumpou I, Arbyn M, Kyrgiou M, Wentzensen N, Koliopoulos G, Martin-Hirsch P, Malamou-Mitsi V, and Paraskevaidis E (2009) p16(INK4a) immunostaining in cytological and histological specimens from the uterine cervix: a systematic review and meta-analysis. Cancer Treat Rev 35**:** 210-20.

van der Aa MA, Schutter EM, Looijen-Salamon M, Martens JE, and Siesling S (2008) Differences in screening history, tumour characteristics and survival between women with screen-detected versus not screen-detected cervical cancer in the east of The Netherlands, 1992–2001. Eur J Obstet Gynecol Reprod Biol 139**:** 204-9.

van Oortmarssen GJ and Habbema JD (1991) Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer 64**:** 559-65.

Viikki M, Pukkala E, and Hakama M (1999) Risk of cervical cancer after a negative Pap smear. J Med Screen 6**:** 103-7.

Viikki M, Pukkala E, and Hakama M (2000) Risk of cervical cancer subsequent to a positive cervical cytology: follow-up study in Finland. Acta Obstet Gynecol Scand 79**:** 576-9.

Virtanen A, Anttila A, Luostarinen T, and Nieminen P (2011) Self-sampling versus reminder letter: effects on cervical cancer screening attendance and coverage in Finland. Int J Cancer 128**:** 2681-7.

von Knebel Doeberitz M, Reuschenbach M, Schmidt D, and Bergeron C (2012) Biomarkers for cervical cancer screening: the role of p16(INK4a) to highlight transforming HPV infections. Expert Rev Proteomics 9**:** 149-63.

Weiss NS (1994) Application of the case-control method in the evaluation of screening. Epidemiol Rev 16**:** 102-8.

Weiss NS (1998) Analysis of case-control studies of the efficacy of screening for cancer: How should we deal with tests done in persons with symptoms? Am J Epidemiol 147**:** 1099-102.

Wilson JM and Jungner YG (1968) Principles and Practice of Screening for Disease. World Health Organization: Geneva.

Wilson SH and Johnson J (1992) An audit of cervical cancer deaths in Nottingham. Cytopathology 3**:** 79-83.

Yajima A, Mori T, Sato S, Wakisaka T, and Suzuki M (1982) Effect of cytologic screening on the detection of cervical carcinoma. Obstet Gynecol 59**:** 565-8.

Yamamoto N, Mori R, Jacklin P, Osuga Y, Kawana K, Shibuya K, and Taketani Y (2012) Introducing HPV vaccine and scaling up screening procedures to prevent deaths from cervical cancer in Japan: a cost-effectiveness analysis. BJOG 119**:** 177-86.

Yang B, Morrell S, Zuo Y, Roder D, Tracey E, and Jelfs P (2008) A case-control study of the protective benefit of cervical screening against invasive cervical cancer in NSW women. Cancer Causes Control 19**:** 569-76.

Ylitalo N, Sørensen P, Josefsson AM, Magnusson PK, Andersen PK, Pontén J, Adami HO, Gyllensten UB, and Melbye M (2000) Consistent high viral load of human papillomavirus 16 and risk of cervical carcinoma in situ: a nested case-control study. Lancet 355**:** 2194-8.

Zapka JG, Taplin SH, Solberg LI, and Manos MM (2003) A framework for improving the quality of cancer care: the case of breast and cervical cancer screening. Cancer Epidemiol Biomarkers Prev 12**:** 4-13.

Zappa M and Ciatto S (2000) Cervix cancer: case-control studies on screening. In *Evaluation and Monitoring of Screening Programmes*, Sankila R, Démaret E, Hakama M, Lynge E, Schouten LJ, and Parkin DM (eds): 99–118. Europe Against Cancer Programme: Luxembourg.

Zappa M, Visioli CB, Ciatto S, Iossa A, Paci E, and Sasieni P (2004) Lower protection of cytological screening for adenocarcinomas and shorter protection for younger women: the results of a case-control study in Florence. Br J Cancer 90**:** 1784-6.

zur Hausen H (1976) Condylomata acuminata and human genital cancer. Cancer Res 36**:** 794.